Annals of the International Society of Dynamic Games

Frank Thuijsman
Florian Wagener
Editors

Advances in Dynamic and Evolutionary Games

Theory, Applications, and Numerical Methods

Birkhäuser

Birkhäuser

Annals of the International Society of Dynamic Games
Volume 14

Series Editor
Tamer Başar

Editorial Board

Tamer Başar, University of Illinois at Urbana-Champaign
Pierre Bernhard, I3S-CNRS and University of Nice-Sophia Antipolis
Maurizio Falcone, University of Rome "La Sapienza"
Jerzy Filar, University of South Australia, Adelaide
Alain Haurie, ORDECSYS, Chêne-Bougeries, Switzerland
Andrzej S. Nowak, University of Zielona Góra
Leon A. Petrosyan, St. Petersburg State University
Alain Rapaport, INRIA, Montpelier
Josef Shinar, Technion, Haifa

More information about this series at http://www.springer.com/series/4919

Frank Thuijsman • Florian Wagener
Editors

Advances in Dynamic and Evolutionary Games

Theory, Applications, and Numerical Methods

Birkhäuser

Editors
Frank Thuijsman
Department of Data Science
and Knowledge Engineering
Maastricht University
Maastricht, The Netherlands

Florian Wagener
Amsterdam School of Economics
Universiteit van Amsterdam
Amsterdam, The Netherlands

Annals of the International Society of Dynamic Games
ISBN 978-3-319-28012-7 ISBN 978-3-319-28014-1 (eBook)
DOI 10.1007/978-3-319-28014-1

Library of Congress Control Number: 2016933763

Mathematics Subject Classification (2010): 91A25, 91A22, 91A23, 91A15, 91A24

© Springer International Publishing Switzerland 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This book is published under the trade name Birkhäuser
The registered company is Springer International Publishing AG Switzerland (www.birkhauser-science.com)

Preface

This volume reflects some of the themes that were discussed during the 2014 Symposium of the International Society of Dynamic Games, which was held 9–12 July 2014 in Amsterdam. The plenary lectures of the Symposium were given by Ariel Pakes, from Harvard, and Eilon Solan, from Tel Aviv University. The tutorial lectures were given by Mark Broom, from London City University, and by Cees Withagen, from the Vrije Universiteit Amsterdam. We are grateful that professors Broom, Pakes and Solan found the opportunity to write a companion piece to their lectures. We are also grateful to professors Altman and Petrosyan, who were both awarded the Isaacs Award for 2015 by the Society, that they contributed to this volume.

In his acceptance speech of the Isaacs Award, Leon Petrosyan reminded the audience that there are still a number of pursuit-evasion games, dating from the early days of differential game theory, that are still poorly understood. In his contribution, he describes several of these: they are united by the common theme that there may be several evaders, or several pursuers. Like in the three-body problem of celestial mechanics, the addition of a third actor in the pursuit-evasion game complicates matters considerably. The remainder of his contribution gives a highly didactical exposition of the theory how cooperative solutions of a dynamic game can be made time consistent by choosing an appropriate distribution of payments, and that this can be supported by a non-cooperative Nash equilibrium in punishment strategies in a specially constructed game. After first illustrating the ideas for the conceptually simpler situation of a game with a finite number of decision nodes, the treatment of cooperative differential games in the last section has by then become fully transparent.

The second recipient of the Isaacs Awards 2015 is Eitan Altman. In his award lecture, he took a bird's-eye perspective and went over many areas of dynamic games that his work has touched upon. For this volume however, he selected just one of these topics. Together with Tania Jiménez, he contributes a joint chapter that revisits the first problem studied in the area of optimal control of queues, which goes back to a seminal paper by Pinhas Naor from 1969. Naor considered problems relating to an M/M/1 queue model, in which decisions have to be made for arrivals on whether or not to enter the queue. Decisions are either made by the individuals themselves or by a social controller. Naor uses full information for the decision makers and derives optimal policies of a threshold type. In their contribution, Altman and Jiménez examine what happens when decision makers are only informed of the queue length being above or below a certain threshold and compare performances with the full information case.

In his contribution, Ariel Pakes investigates the question what kind of dynamic game models can be used to analyse the evolution of competitive industries empirically. He sets out by discussing the common assumption that the behaviour of economic agents or firms can be described by a Markov perfect Nash equilibrium: its limits become evident when trying to apply this to empirical data, as the numerical complexities become formidable. He continues then by the natural question how much can be gained by abandoning the requirement of Markov perfectness while retaining the assumption that firms make rational decisions based on the information that is available to them. This necessitates a rather precise description of the information set and the learning procedure of the firms. He discusses both of these clearly and in great detail, also taking into account the practical issues that arise when developing such a model, and finishes by illustrating it by the maintenance decisions of power-generating firms operating in an electricity market.

In his plenary address, Eilon Solan focused on the theory of zero-sum stochastic games with finitely many states, actions and signals. After defining basic concepts, like λ-discounted value, asymptotic value, uniform value and max-min value, and after surveying the main results on the complete information case, he turned to the incomplete information case and guided the audience to the latest results in this area, as have been derived by Ziliotto. In their joint contribution to this volume, they proceed along the same lines and exhibit the difficulties that arise in case players do not perfectly observe the state or the actions. A relatively simple example based on the Big Match illustrates that the uniform value may well fail to exist. However, it was long thought that every zero-sum stochastic game with signals has an asymptotic value and that this asymptotic value can be characterized by means of the limits of the discounted values and of the finite horizon values. Ziliotto's 2013 example shows that neither of these conjectures is true. In their chapter, Solan and Ziliotti discuss this important example and how it relates to other research questions.

At the Symposium, Mark Broom provided a tutorial on evolutionary game theory that was based on his joint book with Jan Rychtář titled *Game-Theoretical Models in Biology*. Together, they contribute a chapter on the topics that were addressed in the tutorial. Both the static and dynamic aspects of classic evolutionary game theory, and the relationships between them, are explained in a very clear way. Next, the authors explain various ways in which non-linearity can appear in evolutionary games, including pairwise games with strategy-dependent interaction rates and playing the field games, where payoffs depend upon the entire population composition. The chapter also discusses multiplayer games in which the payoffs depend on interactions in groups of more than two players. Altogether, the authors touch upon many topics from evolutionary game theory, give lots of references and provide an excellent introduction to the field.

The contribution by Mahmoud El Chamie and Tamer Başar addresses the problem of designing optimal strategies in consensus protocols for networks that are vulnerable to adversarial attacks. Consensus protocols are based on neighbour-to-neighbour interactions among the nodes in the network. While each node is updating iteratively based on its local information, the goal for the entire set of nodes is to reach consensus. However, the network may be susceptible to attacks from adversaries that want to drive the system away from consensus. The authors propose a game theoretical framework to model this problem as a two-player game between the network designer and the adversary. They use this model to derive optimal strategies for the players within the solution concept of mixed-strategy saddle-point equilibrium. Moreover, they consider a distributed implementation of the optimal control and report on simulations to corroborate the theoretical findings.

Liudmila Kamneva and Valerii Patsko consider a linear differential zero-sum game, where during a fixed time interval two players pull a point over a plane. The first player wins if at the end of the period, the point is located within a given polygonal target set; the second player wins if the point is located in its complement. The force vectors each player can apply are restricted to time-invariant closed convex polygons, possibly degenerate. It is known that if the target set is convex, the appropriate solution set can be constructed exactly in a one-step procedure. In the case that neither the target set nor its complement is convex, Kamneva and Patsko show that the solution set can also be constructed in a one-step procedure, if the time interval is not too large.

In his contribution, Meir Pachter examines a linear-quadratic Gaussian dynamic game (LQGDG) in which the players have partial information. More specifically, the players' initial state and their measurements are private information, but each player is able to observe his opponent's past inputs. It is shown that the specific control-sharing information patterns render the game amenable to solution by the method of dynamic programming, and a Nash equilibrium for the "zero-sum" case of LQGDG is established. Moreover, the correct solution of LQGDGs with a control-sharing information pattern is obtained in closed form.

Josef Shinar, Valery Glizer and Vladimir Turetsky examine a pursuit-evasion differential game of kind with bounded controls and prescribed duration. It is the mathematical model of an interception engagement between two vehicles, a pursuer and an evader, that are both moving with constant velocities in a plane. Each player has two possible dynamics and each can switch between these once during the game. Each player also knows the two possible dynamics the opponent can choose from, but not the actual one that was chosen. The order of the dynamics and the time of change are elements of the player's control that are unknown to the respective opponent, which makes this game a differential game with partial information. An algorithmic example illustrates the complexity of this game with hybrid players.

The chapter by Sourabh Bhattacharya, Ali Khanafer and Tamer Başar relates to the vulnerability of wireless ad hoc networks to security threats. A prominent example of such a threat is jamming, which refers to a malicious attack with the objective to intentionally disrupt communication in the victim network by causing interference or collision of packets. The authors examine a scenario where a team of malicious nodes launches a jamming attack on another team, which is capable of jamming as well. Their analysis takes into consideration constraints on energy and power among the agents. Moreover, they relate the problem of optimal power allocation for communication and jamming to the communication model between the agents. Finally, they provide a sufficient condition for the existence of an optimal decision strategy among the agents based on the physical parameters of the problems.

Sébastien Mitraille and Henry Thille return to an old problem that is still highly relevant and which has already been discussed by Adam Smith in the context of wheat speculation: what is the effect of speculative inventories on prices? In the literature, this question has been extensively discussed for perfectly competitive production, which covers, for instance, many agricultural markets. Mitraille and Thille discuss the situation where there is imperfect competition between producers. Their result confirms Adam Smith's basic insight: speculation effectively helps consumers to smoothen out their consumption over time and always increases consumer's surplus. They also find that it lowers average firm profits. The welfare effects of speculation are ambiguous though: they are positive if there are only a few producers, but negative if there are many.

The contribution by Dharini Hingu, K.S. Mallikarjuna Rao and A.J. Shaiju focuses on evolutionary games with a continuous (pure) strategy space. More specifically, they establish necessary and sufficient conditions for a dimorphic population state P, which is a convex combination of two Dirac measures δ_x and δ_y, to be a rest point of the associated replicator dynamics. They provide necessary and sufficient conditions for the replicator dynamics to converge to P when it starts on the line L between δ_x and δ_y. Moreover, for the special case where all points on L are rest points of the dynamic, they provide sufficient conditions for the dynamics to converge to L, when starting outside L. For the latter case, sufficient conditions to stay away from L are provided as well. Several examples illustrate these results.

Monica Abrudan, Li You, Kateřina Staňková and Frank Thuijsman study which aspects of microbial interactions have an impact of the possible coexistence of different types of bacteria. For this, they use a simple evolutionary game model: bacteria are spread out over a spatial grid, and at each time step, selected cells are attacked by one of their neighbours and possibly replaced by it, depending on the interaction rules. The authors investigated the effects of changing the timing of the interactions, the number of neighbours, the effect of a quorum—attacks are only successful if the attacker is backed up by sufficiently many other neighbours of the same type—and, finally, the mechanics of the interactions. Besides reporting the statistical data obtained by their simulations, they flag an interesting phenomenon, which they call coexistence by small numbers. In that case, a rare species can survive through the protection of a second species with whom it does not interact.

Christopher Andrey, Olivier Bahn and Alain Haurie take the basic idea of robust control theory, to model nature as an antagonistic player, and apply it to a multiplayer game. The translation of robust control to a game setting necessitates some changes in the Nash equilibrium concept, for nature cannot be treated as just another player. Rather each player tries to choose the action optimally, given the actions of the other players, and assuming that nature tries to hit himself hardest. There is some latitude in modelling the impact of nature: here, the authors propose the notion of α-robust equilibrium. The authors propose this as a natural framework to study R&D investments in clean technologies in the face of climate change, and they proceed to give the results of some numerical simulations; in particular, they find that in such a framework, the switch to a low-carbon economy occurs much earlier in the robust equilibrium than in the non-robust one.

The contribution of Frédéric Babonneau, Alain Haurie and Marc Vielle is related to that of Andrey, Bahn and Haurie, in that it also uses a robust Nash equilibrium concept to model an environmental problem, albeit that the concept used here differs slightly from the previous one. Also, their application is different: they investigate pollution abatement and emission permit trading in the face of uncertain abatement costs. Their methods allow them to compute a burden sharing amongst EU countries that distributes equitably over the countries' welfare losses arising from CO_2 abatement costs.

Maastricht, The Netherlands — Frank Thuijsman
Amsterdam, The Netherlands — Florian Wagener
November 2015

Acknowledgements

We would like to thank all authors as well as all reviewers for their willing and prompt support of this volume. We would moreover like to thank GERAD for its continuing financial and administrative support to ISDG activities.

Contents

List of Contributors

Monica Abrudan Faculty of Life Sciences, University of Manchester, Manchester, UK

Eitan Altman INRIA, Sophia Antipolis, France

Christopher Andrey ORDECSYS, Chêne-Bougeries, Switzerland

Frédéric Babonneau ORDECSYS, Chêne-Bougeries, Switzerland

Swiss Federal Institute of Technology at Lausanne (EPFL), Lausanne, Switzerland

Olivier Bahn GERAD and Department of Decision Sciences, HEC Montréal, Montréal, QC, Canada

Tamer Başar Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Sourabh Bhattacharya Department of Mechanical Engineering, Iowa State University, Ames, IA, USA

Mark Broom Department of Mathematics, City University London, London, UK

Mahmoud El Chamie INRIA Sophia Antipolis-Méditerranée, Sophia Antipolis Cedex, France

Valery Y. Glizer Department of Applied Mathematics, Ort Braude College of Engineering, Karmiel, Israel

Alain Haurie ORDECSYS and University of Geneva, Geneva, Switzerland

GERAD, HEC Montréal, Montréal, QC, Canada

Dharini Hingu Department of Mathematics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

Tania Jiménez Avignon University, Avignon, France

Liudmila Kamneva Krasovskii Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia

Institute of Mathematics and Computer Science of Ural Federal University, Ekaterinburg, Russia

Ali Khanafer Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

K.S. Mallikarjuna Rao Industrial Engineering and Operations Research, Indian Institute of Technology Bombay, Mumbai, Maharashtra, India

Sébastien Mitraille Toulouse Business School, University of Toulouse, Toulouse, France

Meir Pachter Department of Electrical and Computer Engineering, Air Force Institute of Technology, AFIT/ENG, Wright-Patterson A.F.B., OH, USA

Ariel Pakes Department of Economics, Harvard University, Cambridge, MA, USA

Valerii Patsko Krasovskii Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia

Institute of Mathematics and Computer Science of Ural Federal University, Ekaterinburg, Russia

Leon A. Petrosyan Faculty of Applied Mathematics, Saint Petersburg State University, Saint Petersburg, Russia

Jan Rychtář Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC, USA

A.J. Shaiju Department of Mathematics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

Josef Shinar Faculty of Aerospace Engineering, Technion - Israel Institute of Technology, Technion City, Haifa, Israel

Eilon Solan School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel

Kateřina Staňková Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht, The Netherlands

Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

Henry Thille Department of Economics & Finance, University of Guelph, Guelph, ON, Canada

Frank Thuijsman Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht, The Netherlands

Vladimir Turetsky Department of Applied Mathematics, Ort Braude College of Engineering, Karmiel, Israel

Marc Vielle Swiss Federal Institute of Technology at Lausanne (EPFL), Lausanne, Switzerland

Li You Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht, The Netherlands

Bruno Ziliotto CEREMADE, Université Paris Dauphine, PSL, Place du Maréchal de Lattre de Tassigny, 75016 Paris, France

Chapter 1
Dynamic Games with Perfect Information

Leon A. Petrosyan

Abstract In this paper we formulate some easy looking but hard to solve problems from pursuit-evasion game theory. Then we focus on the main problem which, from our point of view, arises in dynamic cooperative games: this is the time-inconsistency of optimal solutions. We propose a system of payments, which we call imputation distribution procedure, that can keep the solution time-consistent when the game develops along the cooperative trajectory. It is shown that if payments are made according to an imputation distribution procedure, the cooperative solution can be achieved as a specially constructed Nash equilibrium in punishment strategies. This brings together noncooperative and cooperative approaches in modern game theory, as cooperation can be supported strategically.

Keywords Pursuit-evasion games • Dynamic cooperative games • Imputation distribution procedure

1.1 Introduction

Game theory is one of most sophisticated paradigms that applied mathematics has to offer to study and analyze decision making under real world conditions. Since every human activity is a dynamic process, any type of such activity would be more appropriately analyzed in an inter-temporal framework. One particularly complex and fruitful branch of game theory is the theory of dynamic or differential games, which investigates interactive decision making over time. The dynamic process of pursuit represents a typical conflict and it is therefore not surprising that pursuit and evasion is one of the oldest topics within the field of differential game theory. This area was pioneered by Steinhaus (1925), who was the first to formulate in 1925 the problem of pursuit as differential game of pursuit. After a prolonged silence, in the 1950s mathematicians resumed their research in differential games. The fundamental book of Isaacs (1965) was published and papers of Berkovitz (1964)

L.A. Petrosyan (✉)
Faculty of Applied Mathematics, Saint Petersburg State University, Universitetskiy pr., 35, Petrodvorets, Saint Petersburg 198504, Russia
e-mail: spbuoasis7@peterlink.ru

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_1

and Fleming (1961) appeared. About 10 years later Pontryagin (1967), Krasovskii (1963) and myself (Petrosyan 1965) started to investigate differential games in the USSR. But these were merely zero-sum pursuit-evasion games. In the late 1960s mathematicians turned to n-person differential games (Petrosyan and Murzov 1967). The first papers concerned noncooperative games [see the fundamental book Basar and Olsder (1982)], but starting from the late 1970s cooperative differential game theory was formed.

In this paper we shall formulate some unsolved problems, which we think may encourage specially young mathematicians to get new fruitful results in the theory of differential games and also in practical applications. Of course in a short paper it is not possible to cover the whole variety of unsolved problems, and because of this we shall try to formulate problems which from our point of view are most interesting and about which we were thinking for a long time. We shall start with pursuit games and then will continue with general n-person dynamic and differential games. In the second part of the paper, we concentrate on the time-inconsistency of the main optimality principles in cooperative dynamic game theory.

1.2 Differential Games of Pursuit

In this section we shall deal with pursuit games with simple motion in the plane, since even in the very simple setting many unsolved problems remain. These problems are easy to formulate and to understand but difficult to solve. Their investigation may show the way to the solution of more complicated problems.

Lifeline Games The pursuit takes place in a closed convex set $S \subset R^2$. The equations of motion have the form of "simple motion"

$$
\begin{aligned}
\dot{x}_i &= u_i, \qquad u_1^2 + u_2^2 \le \alpha^2, \\
\dot{y}_i &= v_i, \qquad v_1^2 + v_2^2 \le \beta^2, \\
x_i(t_0) &= \ x_i^0, \quad y_i(t_0) = y_i^0. \\
& i = 1, 2, \quad \alpha > \beta,
\end{aligned} \tag{1.1}
$$

Here $x = (x_1, x_2)$ is the state variable of the Pursuer P and $y = (y_1, y_2)$ that of the Evader E. The payoff function is defined in the following way. Let $x(t)$, $y(t)$ be the solution of (1.1) with initial condition

$$x_0 = (x_1^0, x_2^0), \quad y_0 = (y_1^0, y_2^0) \in S.$$

Let moreover

$$t_{S_P} = \inf\{t : x(t) \in S\}, \quad t_{S_E} = \inf\{t : y(t) \in S\}.$$

It is supposed that $x(t) \notin S$ if $t \ge t_{S_P}$ and $y(t) \notin S$ if $t \ge t_{S_E}$. Let finally

$$t_P = \min\{t : \rho(x(t), y(t)) = l; \quad l \ge 0\},$$

where $\rho(x, y)$ is the Euclidean distance.

The game is zero-sum, with the payoff K of player P defined as

$$K(x_0, y_0; u(\cdot), v(\cdot)) = \begin{cases} +\,1, & \text{if} \quad t_p \leq t_{S_E}, \\ \quad 0, & \text{if} \quad t_p = t_{S_E} = \infty, \\ -\,1, & \text{if} \quad t_p > t_{S_E}. \end{cases}$$

It is supposed that at each instant of time player P knows t, $x(t)$, $y(t)$, $v(t)$, and E knows t, $x(t)$, $y(t)$.

The solution of this game is known (see Petrosyan 1965). For $l = 0$ it can be described geometrically as follows. The optimal strategy of P is the parallel pursuit strategy (Π-strategy): P has to approach E with maximal velocity in such a way that the segment $[x(t), y(t)]$ moves parallel to itself (i.e. parallel to $[x_0, y_0]$). The optimal strategy of E is to move with maximal velocity along a straight line on which her capture occurs after she penetrates the boundary of S. If such a straight line does not exist, any strategy of E is optimal and E will be captured in S. The solution in the case $l > 0$ is similar (Dutkevich and Petrosyan 1972).

Suppose now that we have two evaders E_1, E_2, and that P can l-capture (to approach up to distance $l > 0$) both of them if playing against either E_1 or E_2 separately. But what if E_1, E_2 cooperate and want to maximize the number of alive evaders, that is, the number of evaders that reach the boundary of S before capture (this number can be only 1 or 0)? The solution is unknown.

What is the difficulty? Suppose P is using the Π-strategy against E_1; then the question is what E_2 should do in the time interval $[t_0, t_{P_1}]$, where t_{P_1} is the capture time of the first evader E_1, and how E_1 should behave to maximize the time t_{P_1} and to bring the capture point in the worst position for P to continue the pursuit against E_2. This problem is not solved yet.

Similarly there is no complete solution of the simple pursuit game in the plane with one pursuer P and m evaders $E_1, \ldots, E_m$. In the case $m = 2$ the solution was found in Petrosyan and Shiryaev (1978) for the case when the pursuer is restricted to choose one of the following two strategies: to use the parallel pursuit strategy (Π-strategy) to pursue E_1 first, and E_2 afterwards, or conversely to pursue E_2 first and E_1 afterwards (see Petrosyan and Shiryaev 1978). More general cases are however still open.

It is also interesting to consider the (noncooperative) non-zero sum game between $E_1, \ldots, E_m$ and P, where each of evaders plays for herself. For some initial conditions and some parameter values it is possible to construct a very strange Nash equilibrium, where P dictates the behavior of the evaders by using the threat to start the pursuit of those evaders which do not follow his orders (and these will be captured first).

Also not solved is the pursuit evasion game with a team of pursuers and one evader in the following case: the pursuers move faster than the evader ($\alpha > \beta$), the payoff is the capture time, and at the beginning the evader is surrounded by the team of pursuers.

1.3 Cooperative Multistage Games with Perfect Information

In what follows we shall consider a game in extensive form with perfect information as basic model.

Definition 1. A game tree is a finite oriented treelike graph Y with root x_0.

We shall use the following notations. Let x be some vertex (position). We denote by $Y(x)$ a subtree Y with root in x, and by $Z(x)$ the immediate successors of x. Vertices y that directly follow after x are called alternatives in x ($y \in Z(x)$). The player who makes a decision in x, and thereby selects the next alterative position in x, will be denoted by $i(x)$. The choice of player $i(x)$ in position x will be denoted by $\bar{x} \in Z(x)$.

Let $N = \{1, \cdots, n\}$ be the set of all players in the game.

Definition 2. A game in extensive form with perfect information $G(x_0)$ is a graph tree $Y(x_0)$ with the following additional properties:

1. The set of vertices (positions) is split into $n + 1$ subsets $P_1, P_2, \ldots, P_{n+1}$ that form a partition of the set of all vertices of the graph tree Y. For $i = 1, \ldots, n$, the vertices (positions) $x \in P_i$ are called the personal positions of player i; vertices (positions) $x \in P_{n+1}$ are called terminal positions.
2. Each pair of vertices $x \notin P_{n+1}$ and $y \in Z(x)$ defines an arc (x, y) of the graph Y. On each arc (x, y), there are n real numbers $h_i(x, y)$, $i = 1, \ldots, n$ defined, the payoffs of players on this arc. Moreover, for $x \in P_{n+1}$, $i = 1, \ldots, n$, terminal payoffs $H_i(x)$ of player i are given.

See Kuhn (1953) for this definition.

Definition 3. A strategy of player i is a mapping $U_i(\cdot)$ that associates to each position $x \in P_i$ a unique alternative $y \in Z(x)$.

Denote by $K_i(x; u_1(\cdot), \ldots, u_n(\cdot))$ the payoff function of player $i \in N$ in the subgame $G(x)$ starting from the position x, when the n-tuple of strategies $(u_1(\cdot), \ldots, u_n(\cdot))$ is played. That is

$$K_i(x; u_1(\cdot), \ldots, u_n(\cdot)) = \sum_{k=0}^{l-1} h_i(x_k, x_{k+1}) + H_i(x_l),$$

where $x_l \in P_{n+1}$ is the last vertex (position) in the path $\tilde{x} = (x_0, x_1, \ldots, x_l)$ realized in the subgame $G(x)$.

Furthermore, denote by $\bar{u}(\cdot) = (\bar{u}_1(\cdot), \ldots, \bar{u}_n(\cdot))$ the n-tuple of strategies, and by $\bar{x} = (\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_l)$, $\bar{x}_l \in P_{n+1}$ the trajectory (path), which satisfy

$$\max_{u_1(\cdot),\ldots,u_n(\cdot)} \sum_{i=1}^{n} K_i(x_0; u_1(\cdot), \ldots, u_n(\cdot))$$
$$= \sum_{i=1}^{n} K_i(x_0; \bar{u}_1(\cdot), \ldots, \bar{u}_n(\cdot)) = \sum_{i=1}^{n} \left(\sum_{k=0}^{l-1} h_i(\bar{x}_k, \bar{x}_{k+1}) + H_i(\bar{x}_l) \right). \qquad (1.2)$$

This path $\bar{x}$ we shall call "optimal cooperative trajectory".

The characteristic function of $G(x_0)$ is defined classically by

$$V(x_0;N) = \sum_{i=1}^{n}\left(\sum_{k=0}^{l-1} h_i(\bar{x}_k,\bar{x}_{k+1}) + H_i(\bar{x}_l)\right),$$
$$V(x_0;\emptyset) = 0, \qquad V(x_0;S) = \operatorname{Val} G_{S,N\setminus S}(x_0),$$

where $\operatorname{Val} G_{S,N\setminus S}(x_0)$ is the value of the zero-sum game that is played between the coalition S acting as first player and the coalition $N \setminus S$ acting as second player, with the payoff of player S equal to

$$\sum_{i\in S} K_i(x_0;u_1(\cdot),\ldots,u_n(\cdot)).$$

Using the characteristic function we define the set $C(x_0)$ of imputations in the game $G(x_0)$ as

$$C(x_0) = \left\{ \xi = (\xi_1,\ldots,\xi_n) \quad : \quad \xi_i \geq V(x_0;\{i\}), \quad \sum_{i\in N}\xi_i = V(x_0;N)\right\},$$

and the core $M_{\text{core}}(x_0) \subset C(x_0)$ as

$$M_{\text{core}}(x_0) = \left\{ \xi = (\xi_1,\ldots,\xi_n) \quad : \quad \sum_{i\in S}\xi_i \geq V(x_0;S), \quad S \subset N\right\}.$$

Other optimality principles of classical game theory, as Neumann–Morgenstern (NM) solutions, Shapley values etc. are defined analogously. In what follows we shall denote by $M(x_0) \subset C(x_0)$ any of these optimality principles.

Suppose that at the beginning of the game players agree to use the optimality principle $M(x_0) \subset C(x_0)$ as basis for the selection of the "optimal" imputation $\bar{\xi} \in M(x_0)$. Playing cooperatively means then that after choosing a common strategy maximizing the common payoff, each player obtains a payoff $\bar{\xi}_i$ from the optimal imputation $\bar{\xi} \in M(x_0)$ after the end of the game, and after the maximal common payoff $V(x_0;N)$ is really earned by the players.

But when the game G actually develops along the "optimal" trajectory $\bar{x} = (\bar{x}_0,\bar{x}_1,\ldots,\bar{x}_l)$, at each vertex $\bar{x}_k$ the players find themselves in a new multistage game with perfect information $G(\bar{x}_k)$, $k = 0,\ldots,l$, which is the subgame of the original game G that starts at $\bar{x}_k$, and which has payoffs

$$K_i(\bar{x}_k;u_1(\cdot),\ldots,u_n(\cdot)) = \sum_{j=k}^{l-1} h_i(x_j,x_{j+1}) + H_i(x_l), \quad i = 1,\ldots,n.$$

It is important to mention that for problem (1.2) the Bellman optimality principle holds: the part $\bar{x}^k = (\bar{x}_k, \ldots, \bar{x}_j, \ldots, \bar{x}_l)$ of the trajectory $\bar{x}$ that starts at $\bar{x}_k$ maximizes the sum of the payoffs in the subgame $G(\bar{x}_k)$, i.e.

$$\max_{x_k,\ldots,x_j,\ldots,x_l} \sum_{i=1}^{n} \left[\sum_{j=k}^{l-1} h_i(x_j, x_{j+1}) + H_i(x_l) \right] = \sum_{i=1}^{n} \left[\sum_{j=k}^{l-1} h_i(\bar{x}_j, \bar{x}_{j+1}) + H_i(\bar{x}_l) \right],$$

which means that the trajectory $\bar{x}^k = (\bar{x}_k, \ldots, \bar{x}_j, \ldots, \bar{x}_l)$ is also "optimal" in the subgame $G(\bar{x}_k)$.

Before entering the subgame $G(\bar{x}_k)$, player i has already earned the amount

$$K_i^{\bar{x}_k} = \sum_{j=0}^{k-1} h_i(\bar{x}_j, \bar{x}_{j+1}), \quad i = 1, \ldots, n.$$

Moreover, at the beginning of the game $G = G(x_0)$ player i was expecting to get the payoff $\bar{\xi}_i$—the ith component of the "optimal" imputation $\bar{\xi} \in M(x_0) \subset C(x_0)$. From this it follows that in the subgame $G(\bar{x}_k)$ he expects to get a payoff equal to

$$\bar{\xi}_i - K_i^{\bar{x}_k} = \bar{\xi}_i^{\bar{x}_k}, \quad i = 1, \ldots, n.$$

Then the question arises whether the new vector

$$\bar{\xi}^{\bar{x}_k} = \left(\bar{\xi}_1^{\bar{x}_k}, \ldots, \bar{\xi}_i^{\bar{x}_k}, \ldots, \bar{\xi}_n^{\bar{x}_k} \right)$$

is in the same sense optimal for the subgame $G(\bar{x}_k)$ as the vector $\bar{\xi}$ is for the game $G(\bar{x}_0)$. If this is not the case, it means that the players in the subgame $G(\bar{x}_k)$ will not orient themselves on the same optimality principle as for the game $G(\bar{x}_0)$. This may lead them to end the cooperation by changing the chosen cooperative strategies $\bar{u}_i(\cdot)$ and thus changing the optimal trajectory $\bar{x}$ in the subgame $G(\bar{x}_k)$. We shall now try to formalize this reasoning.

Introduce for each subgame $G(\bar{x}_k)$, $k = 1, \ldots, l$, the characteristic function $V(\bar{x}_k; S)$, $S \subset N$, in the same manner as it was done for the game $G(x_0)$. As before, but now based on the characteristic function $V(\bar{x}_k; S)$, we introduce the set of imputations

$$C(\bar{x}_k) = \left\{ \xi = (\xi_1, \ldots, \xi_n) \quad : \quad \xi_i \geq V(\bar{x}_k; \{i\}), \quad \sum_{i \in N} \xi_i = V(\bar{x}_k; N) \right\},$$

the core $M_{\text{core}}(\bar{x}_k) \subset C(\bar{x}_k)$

$$M_{\text{core}}(\bar{x}_k) = \left\{ \xi = (\xi_1, \ldots, \xi_n) \quad : \quad \sum_{i \in S} \xi_i \geq V(\bar{x}_k; S), \quad S \subset N \right\},$$

NM solutions, Shapley values and all the other standard solution concepts. Denote again by $M(\bar{x}_k) \subset C(\bar{x}_k)$ the optimality principle that is used in the subgame $G(\bar{x}_k)$. This is the same optimality principle that was selected by the players in the game $G(x_0)$.

If we suppose that the players of the game $G(x_0)$, when moving along the optimal trajectory $(\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_l)$, stick to the same optimality principle M, then the vector $\bar{\xi}^{\bar{x}_k} = \bar{\xi} - K^{\bar{x}_k}$ must belong to the set $M(\bar{x}_k)$ for all $k = 0, \ldots, l$.

It is very difficult to find games and corresponding solution concepts for which this condition is satisfied. We shall illustrate this using the following example.

Suppose that for a game G, we have that $h_i(x_k, x_{k+1}) = 0$ for $k = 0, \ldots, l-1$, and $H_i(x_l) \neq 0$; that is, the game G is a game with terminal payoff only. Then the last condition would mean that

$$\bar{\xi} = \bar{\xi}^{\bar{x}_k} \in M(\bar{x}_k), \quad k = 0, \ldots, l, \tag{1.3}$$

which implies

$$\bar{\xi} \in \bigcap_{k=0}^{l} M(\bar{x}_k). \tag{1.4}$$

For $k = l$, it follows from Eq. (1.3) that

$$\bar{\xi} \in M(\bar{x}_l).$$

But $M(\bar{x}_l) = C(\bar{x}_l) = \{H_i(\bar{x}_l)\}$. Moreover, this has to be valid for all imputations of the set $M(\bar{x}_0)$ and for all optimality principles $M(x_0) \subset C(x_0)$, which means that in the cooperative game with terminal payoffs the only reasonable optimality principle is

$$\bar{\xi} = \{H_i(\bar{x}_l)\},$$

the payoff vector obtained at the endpoint of the cooperative trajectory in the game $G(x_0)$. At the same time, simple examples show that the intersection (1.4), except for trivial cases, is void for games with terminal payoffs.

How to overcome this difficulty? A plausible way is to introduce a special payment rule, a stage salary, for each stage of the game in such a way that the following two conditions are met. First, the payments at each stage do not exceed the common amount earned by the players until this stage; second, the payments received by the players starting from the stage k, in the subgame $G(\bar{x}_k)$, belong to the same optimality principle as the imputation $\bar{\xi}$ on which players agree in the game $G(x_0)$ at the beginning of the game. We shall now investigate whether or not it is possible to find such a payment rule.

We introduce the notion of imputation distribution procedure (IDP).

Definition 4. Suppose that $\xi = \{\xi_1, \dots, \xi_i, \dots, \xi_n\} \in M(x_0)$. Any matrix $\beta = \{\beta_{ik}\}$, with $i = 1, \dots, n$, $k = 0, \dots, l$, such that $\beta_{ik} \geq 0$ for all i and k, and such that

$$\xi_i = \sum_{k=0}^{l} \beta_{ik}, \tag{1.5}$$

is called an imputation distribution procedure (IDP).

Denote $\beta_k = (\beta_{1k}, \dots, \beta_{nk})$ and $\beta(k) = \sum_{m=0}^{k-1} \beta_m$. The interpretation of the IDP β is that β_{ik} is the payment to player i at stage k of the game $G(x_0)$, i.e. at the first stage of the subgame $G(\bar{x}_k)$. From Definition 4 it follows that in the game $G(x_0)$ each player i gets the amount ξ_i, $i = 1, \dots, n$, which he expects to get as the ith component of the optimal imputation $\xi_i \in M(x_0)$ in the game $G(x_0)$.

Likewise, the quantity $\beta_i(k)$ is the amount received by player i after the first k stages of the game $G(x_0)$.

Definition 5. The optimality principle $M(x_0)$ is called time-consistent if for every $\xi \in M(x_0)$ there exists an IDP β such that

$$\xi^k = \xi - \beta(k) \in M(\bar{x}_k), \quad k = 0, 1, \dots, l. \tag{1.6}$$

Definition 6. The optimality principle $M(x_0)$ is called strongly time-consistent if for every $\xi \in M(x_0)$ there exists an IDP β such that

$$\beta(k) \oplus M(\bar{x}_k) \subset M(x_0), \quad k = 0, 1, \dots, l;$$

here $a \oplus A = \{a + a' : \; a' \in A, \, a \in R^n, \, A \subset R^n\}$.

Time-consistency of the optimality principle $M(x_0)$ implies that for each imputation $\xi \in M$ there exits an IDP β such that the following holds: if on each arc $(\bar{x}_k, \bar{x}_{k+1})$ of the optimal trajectory $\bar{x}$ payments are made to the players according to β, then in every subgame $G(\bar{x}_k)$ the players may expect to receive payments $\bar{\xi}^k$ that are optimal in the subgame $G(\bar{x}_k)$ in the same sense as in the game $G(x_0)$.

Strong time-consistency means the following: assume that payments are made in stage k according to an IDP β, and that the players then reconsider the choice of (optimal) imputation in the subgame. If they stick to the same optimality principle in the subgame $G(\bar{x}_k)$ as they use in the game $G(x_0)$, then the resulting modified imputation for the game $G(x_0)$ still belongs to the set $M(x_0)$ of imputations that are optimal under the optimality principle M.

For any optimality principle $M(x_0) \subset C(x_0)$ and for every imputation $\bar{\xi} \in M(x_0)$ we can define β_{ik} by the following formulas

$$\begin{aligned} \beta_{ik} &= \bar{\xi}_i^{\bar{x}_k} - \bar{\xi}_i^{\bar{x}_{k+1}}, \quad i = 1, \dots, n, \quad k = 0, \dots, l-1 \\ \beta_{il} &= \bar{\xi}_i^{\bar{x}_l}, \qquad i = 1, \dots, n. \end{aligned} \tag{1.7}$$

It follows that

$$\sum_{k=0}^{l} \beta_{ik} = \sum_{k=0}^{l-1} \left(\bar{\xi}_i^{\bar{x}_k} - \bar{\xi}_i^{\bar{x}_k+1} \right) + \bar{\xi}_i^{\bar{x}_l} = \bar{\xi}_i^{\bar{x}_0} = \bar{\xi}_i,$$

as well as

$$\bar{\xi} - \beta(k) = \bar{\xi}^{\bar{x}_k} \in M(\bar{x}_k), \quad k = 0, \ldots, l.$$

This last inclusion implies time consistency of $M(x_0)$.

The strong time consistency condition is more involved: for instance, we cannot even derive a formula like (1.7).

The Regularized Game G_α For every $\alpha \in M(x_0)$ we introduce a noncooperative game $G_\alpha(x_0)$, the regularization of $G(x_0)$, that differs from the game $G(x_0)$ only by having different payoffs along optimal cooperative path $\bar{x} = (\bar{x}_0, \ldots, \bar{x}_l)$. It is possible to choose G_α such that there exists a Nash equilibrium in $G_\alpha(x_0)$ with payoffs equal to α.

To see this, take $\alpha \in M(x_0)$. Define an imputation distribution procedure (IDP), given by $\beta(k) = (\beta_{1k}, \ldots, \beta_{nk})$ for $k = 0, 1, \ldots, l$, such that

$$\alpha_i = \sum_{k=0}^{l} \beta_{ik}. \tag{1.8}$$

Let the subgames $G(\bar{x}_k)$ along the optimal path $\bar{x}$, as well as the imputation sets $C(\bar{x}_k)$, be defined as before. Denote by $K_i^\alpha(x_0; u_1(\cdot), \ldots, u_n(\cdot))$ the payoff function in the game $G_\alpha(x_0)$.

Suppose $x = (x_1, x_2, \ldots, x_l)$ is the path resulting from the initial state x_0, when strategies $(u_1(\cdot), \ldots, u_n(\cdot))$ are played, and suppose that m is the maximal index for which $x_k = \bar{x}_k$, that is, the maximal number of stages in which the path coincides with the cooperative path $\bar{x}$. Then

$$K_i^\alpha(x_0; u_1(\cdot), \ldots, u_n(\cdot)) = \sum_{k=0}^{m-1} \beta_{ik} + \sum_{k=m}^{l-1} h_i(x_k, x_k + 1) + H_i(x_l)$$

and

$$K_i^\alpha(x_0; \bar{u}_1(\cdot), \ldots, \bar{u}_n(\cdot)) = \alpha_i.$$

By defining the payoff function in the game $G_\alpha(x_0)$ appropriately, we obtain that the payoffs along the optimal cooperative trajectory are equal to the components of the imputation $\alpha = (\alpha_1, \ldots, \alpha_n)$.

Definition 7. The game $G_\alpha(x_0)$ is called a regularization (or α-regularization) of the game $G(x_0)$ if an IDP β can be defined in such a way that

$$\alpha_i^k = \sum_{j=k}^{l} \beta_{ij}.$$

Note: this implies that

$$\beta_{ik} = \alpha_i^k - \alpha_i^{k+1}$$

for all $i = 1, \ldots, n$ and all $k = 0, 1, \ldots, l-1$, and

$$\beta_{il} = \alpha_i^l \quad \text{and} \quad \alpha_i = \alpha_i^0$$

for all i. Moreover $\alpha^k \in C(\bar{x}_k)$ for all k.

Theorem 1. *In the regularization $G_\alpha(x_0)$ of the game $G(x_0)$ there exist a Nash equilibrium with payoffs $\alpha = (\alpha_1, \ldots, \alpha_n)$.*

Proof. Along the cooperative path we have

$$\alpha_i^k \geq V(\bar{x}_k; \{i\}), \quad i = 1, \ldots, n, \quad k = 0, 1, \ldots, l,$$

since $\alpha^k = (\alpha_1^k, \ldots, \alpha_n^k) \in C(\bar{x}_k)$ is an imputation in $G(\bar{x}_k)$; note that $V(\bar{x}_k; \{i\})$ is computed for the subgame $G(\bar{x}_k)$, not for $G_\alpha(\bar{x}_k)$. At the same time

$$\alpha_i^k = \sum_{j=k}^{l} \beta_{ij}$$

and we obtain

$$\sum_{j=k}^{l} \beta_{ij} \geq V(\bar{x}_k; \{i\}), \quad i = 1, \ldots, n, \quad k = 0, 1, \ldots, l. \tag{1.9}$$

But $\sum_{j=k}^{l} \beta_{ij}$ is the payoff of player i in the subgame $G_\alpha(\bar{x}_k)$ along the cooperative path, and from (1.9) one can construct a Nash equilibrium in punishment strategies with payoffs $\alpha = (\alpha_1, \ldots, \alpha_n)$ and resulting cooperative path $\bar{x} = (\bar{x}_0, \ldots, \bar{x}_l)$.

Example. In this example we shall take the Shapley value (Shapley 1953) as solution concept. Using the proposed regularization of the game, we shall see that there exists a Nash equilibrium with payoffs equal to the components of the Shapley value.

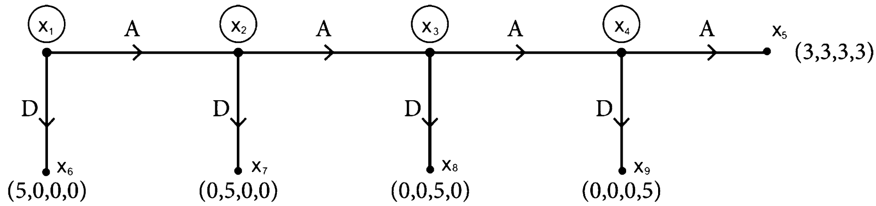

Fig. 1.1 Game $G(x_1)$

Table 1.1 Characteristic functions of the games $G(x_k)$, for $k = 1, \ldots, 5$

i	1	2	3	4	5
$V_i(1,2,3,4)$	12	12	12	12	12
$V_i(1,2,3)$	5	5	5	0	9
$V_i(1,2,4)$	5	5	0	9	9
$V_i(1,3,4)$	5	0	9	9	9
$V_i(2,3,4)$	0	9	9	9	9
$V_i(1,2)$	5	5	0	0	6
$V_i(1,3)$	5	0	5	0	6
$V_i(1,4)$	5	0	0	6	6
$V_i(2,3)$	0	5	5	0	6
$V_i(2,4)$	0	5	0	6	6
$V_i(3,4)$	0	0	6	6	6
$V_i(1)$	5	0	0	0	3
$V_i(2)$	0	5	0	0	3
$V_i(3)$	0	0	5	0	3
$V_i(4)$	0	0	0	5	3

There are four players ($n = 4$) and nine vertices $x_1, \ldots, x_9$. The sets of personal positions of the players are $P_1 = \{x_1\}$, $P_2 = \{x_2\}$, $P_3 = \{x_3\}$, and $P_4 = \{x_4\}$. The set of terminal positions is $P_5 = \{x_5, x_6, x_7, x_8, x_9\}$. There are only terminal payoffs; that is $h_i(x_k, x_{k+1}) = 0$ for all i and all k. The terminal payoffs are $g(x_5) = (3,3,3,3)$, $g(x_6) = (5,0,0,0)$, $g(x_7) = (0,5,0,0)$, $g(x_8) = (0,0,5,0)$, and $g(x_9) = (0,0,0,5)$. See Fig. 1.1 for the game $G(x_1)$.

The cooperation path is clearly

$$\bar{x} = \{x_1, x_2, x_3, x_4, x_5\}.$$

It can be easily seen that (D, D, D, D) is a Nash equilibrium, and that the cooperative path (A, A, A, A) is not a Nash equilibrium.

The characteristic function of the game $G(x_1)$ is given in the first column of Table 1.1. Figure 1.2 gives the subgames $G(x_k)$ of $G(x_1)$, for $k = 2, \ldots, 5$; their characteristic functions are also given in Table 1.1.

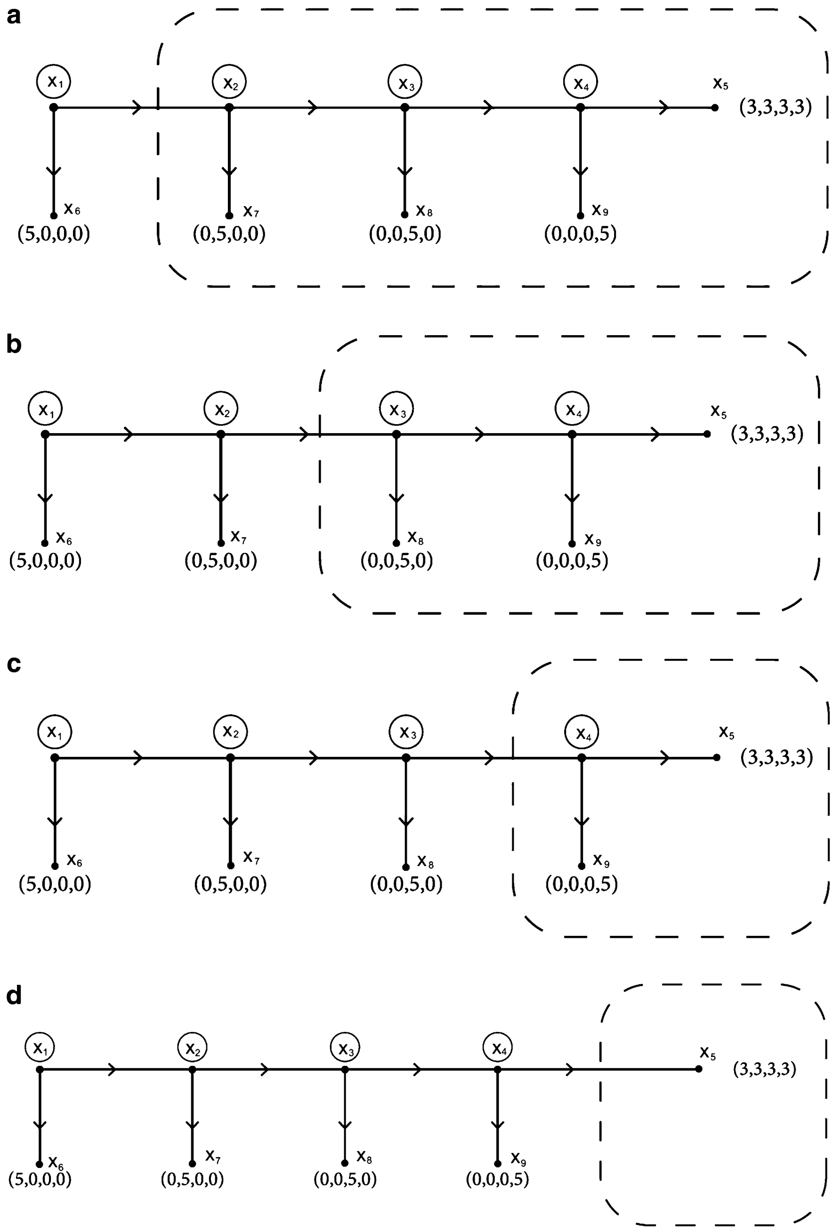

Fig. 1.2 Subgames of $G(x_1)$. (**a**) Subgame $G(x_2)$. (**b**) Subgame $G(x_3)$. (**c**) Subgame $G(x_4)$. (**d**) Subgame $G(x_5)$

The Shapley values Sh^k of $G(x_k)$ are, respectively:

$$Sh^1 = \left(\frac{27}{4}, \frac{7}{4}, \frac{7}{4}, \frac{7}{4}\right);$$

$$Sh^2 = \left(\frac{9}{12}, \frac{85}{12}, \frac{25}{12}, \frac{25}{12}\right),$$

and $Sh^1 \neq Sh^2$;

$$Sh^3 = \left(1, 1, \frac{90}{12}, \frac{30}{12}\right),$$

and $Sh^2 \neq Sh^3$;

$$Sh^4 = \left(\frac{16}{12}, \frac{16}{12}, \frac{16}{12}, \frac{96}{12}\right);$$

and $Sh^4 \neq Sh^3$;

$$Sh^5 = (3, 3, 3, 3)\,,$$

and $Sh^5 \neq Sh^4$.

Compute now the IDP (imputation distribution procedure): for $k = 1, \ldots, 4$:

$$Sh^k = \beta(k) + Sh^{k+1},$$

implying

$$\beta(k) = Sh^k - Sh^{k+1}$$

or

$$\sum_{k=j}^{5} \beta(k) = Sh^j$$

for $j = 1, \ldots, 5$. This results in

$$\beta(1) = \left(\frac{72}{12}, -\frac{64}{12}, -\frac{4}{12}, -\frac{4}{12}\right),$$

$$\beta(2) = \left(\frac{-3}{12}, \frac{73}{12}, -\frac{65}{12}, \frac{-5}{12}\right),$$

$$\beta(3) = \left(-\frac{4}{12}, -\frac{4}{12}, \frac{74}{12}, -\frac{66}{12}\right),$$

$$\beta(4) = \left(-\frac{20}{12}, -\frac{20}{12}, -\frac{20}{12}, \frac{60}{12}\right),$$

$$\beta(5) = (3, 3, 3, 3)\,.$$

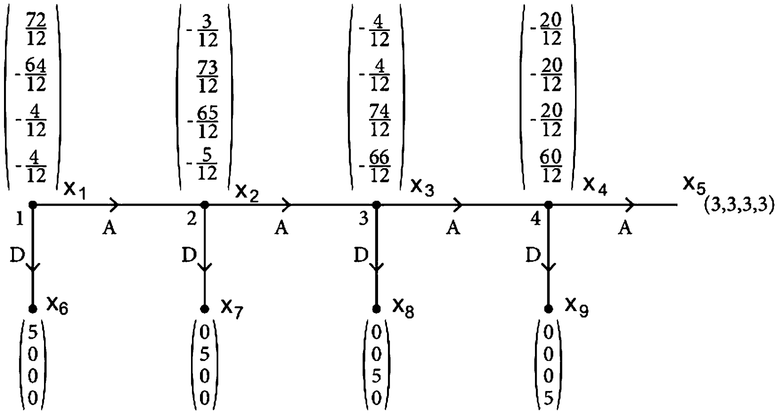

Fig. 1.3 Game G_α

The regularization G_α of the game $G(x_1)$ [when α is the Shapley value in $G(x_1)$] and the Nash equilibrium strategically supported cooperation are given in Fig. 1.3.

The payoffs

$$\left(\frac{72}{12}, -\frac{64}{12}, -\frac{4}{12}, -\frac{4}{12}\right), \left(\frac{-3}{12}, \frac{73}{12}, -\frac{65}{12}, \frac{-5}{12}\right),$$
$$\left(-\frac{4}{12}, -\frac{-4}{12}, \frac{74}{12}, -\frac{-66}{12}\right), \left(-\frac{20}{12}, -\frac{20}{12}, -\frac{20}{12}, \frac{60}{12}\right)$$

are defined on the arcs $(1,2)$, $(2,3)$, $(3,4)$, $(4,5)$ respectively.

We can see that the inequalities (1.9) hold for the game G_α.

$$\sum_{j=1}^{4} \beta_{1j} = \frac{72}{15} + \frac{-3}{12} - \frac{4}{12} - \frac{20}{12} + 3 = \frac{27}{4} > 5 = V(\bar{x}_1; \{1\}),$$

$$\sum_{j=2}^{4} \beta_{2j} = \frac{73}{12} - \frac{4}{12} - \frac{20}{12} + 3 = \frac{85}{12} > 5 = V(\bar{x}_2; \{2\}),$$

$$\sum_{j=2}^{4} \beta_{3j} = \frac{74}{12} - \frac{20}{12} + 3 = \frac{15}{2} > 5 = V(\bar{x}_3; \{3\}),$$

$$\sum_{j=3}^{4} \beta_{4j} = \frac{60}{12} + 3 = \frac{32}{4} > 5 = V(\bar{x}_4; \{4\}).$$

This means that in the Nash equilibrium (A, A, A, A) of the regularized game G_α, the payoffs are $(\frac{27}{4}, \frac{7}{4}, \frac{7}{4}, \frac{7}{4})$, and hence exactly equal to the Shapley value. Thus the computed Nash equilibrium supports the cooperative outcome and the cooperative payoffs of the original game; these payoffs are redistributed according to an IDP that guarantees the time-consistency of the Shapley value.

1.4 Cooperative Differential Games

In this section we extend the ideas introduced in the previous section to differential games. We define cooperative differential games in characteristic function form, and introduce the notions of optimality principle and, based on it, solution concept. "Imputation distribution procedures" (IDP) are defined for the differential case and connected to the basic definitions of time-consistency and strongly time-consistency. Finally, we derive sufficient conditions of the existence of time consistent solutions.

1.4.1 Definition of Cooperative Differential Games in Characteristic Function Form

We will investigate n-person differential games starting from an initial state $x_0 \in R^n$ at an initial time $t_0 \in R^1$ and with a prescribed duration $T - t_0$ where the end time $T > t_0$ is a finite number. To indicate the dependence of the game on initial state and duration, we denote it by $\Gamma(x_0, T - t_0)$.

Let $N = \{1, \ldots, n\}$ be the set of all players of the game, and let the equations of motion have the form

$$\dot{x} = f(x, u_1, \ldots, u_n), \quad x(t_0) = x_0; \tag{1.10}$$

here $x \in R^n$ is the state variable and $u_i \in U_i$ is the control variable of player $i \in N$, with U_i a compact set. The payoff function of player i is defined in the following way:

$$K_i(x_0, T - t_0; u_1, \ldots, u_n) = \int_{t_0}^{T} h_i(x(t))dt + H_i(x(T)), \tag{1.11}$$

where $h_i(x) > 0$, $H_i(x) > 0$ are given positive continuous functions and where $x(t)$ is the trajectory realized from the initial state x_0 under the strategy choice $u = (u_1, \ldots, u_n)$ of the players. We restrict attention to feedback or closed loop strategies $u_i = u_i(t, x)$, $t \in [t_0, T]$, $x \in R^n$.

Consider the cooperative form of the game $\Gamma(x_0, T - t_0)$. That is, we suppose that the players, before starting the game, agree to play strategies $u_1^*, \ldots, u_n^*$ such that the resulting trajectory $x^*(t)$ maximizes the sum of the payoffs

$$\max_u \sum_{i=1}^{n} K_i(x_0, T - t_0; u_1, \ldots, u_n) = \sum_{i=1}^{n} K_i(x_0, T - t_0; u_1^*, \ldots, u_n^*)$$
$$= \sum_{i=1}^{n} \left[\int_{t_0}^{T} h_i(x^*(t))dt + H_i(x^*(T)) \right] = v(N; x_0, T - t_0),$$

The trajectory $x^*(t)$ is called conditionally optimal.

To define the cooperative game one has to introduce the characteristic function. We will do this in a classical fashion by considering a zero-sum game, defined with the same structure as the game $\Gamma(x_0, T - t_0)$, between the coalition S as first player and the coalition $N \backslash S$ as second player, where the payoff of S is equal to the sum of payoffs of players from S. Denote this game as $\Gamma_S(x_0, T - t_0)$. Suppose that the value $v(S; x_0, T - t_0)$ of this game exists (existence of values of zero-sum differential games has been proved under very general conditions). The characteristic function is then defined for each $S \subset N$ as the value $v(S; x_0, T - t_0)$ of $\Gamma_S(x_0, T - t_0)$.

Note that the positiveness of the payoff functions K_i, $i = 1, \ldots, n$ implies positiveness of the characteristic function. From the superadditivity of v it follows moreover that

$$v(S'; x_0, T - t_0) \geq v(S; x_0, T - t_0)$$

for all coalitions $S, S' \subset N$ such that $S \subset S'$; that is, superadditivity of v in S implies that v is monotone in S.

The pair $(N, v(S; x_0, T - t_0))$, where N is the set of players and v the characteristic function, is called a cooperative differential game in characteristic function form, and is denoted by $\Gamma_v(x_0, T - t_0)$.

A method to allocate the total profit "equitably" among the players constitutes a solution of the cooperative game. The set of equitable allocations satisfying an optimality principle is a solution of the cooperative game in the sense of this optimality principle.

We will now define solutions of the game $\Gamma_v(N; x_0, T - t_0)$.

Definition 8. A vector $\xi = (\xi_1, \ldots, \xi_n)$, whose components satisfy the conditions:

(1) $\xi_i \geq v(\{i\}; x_0, T - t_0)$ for each $i \in N$;
(2) $\sum_{i \in N} \xi_i = v(N; x_0, T - t_0)$,

is called an imputation in the game $\Gamma_v(x_0, T - t_0)$.

Here ξ_i should be thought of as the share of player $i \in N$ in the total gain $v(N; x_0, T - t_0)$.

Denote the set of all imputations in $\Gamma_v(x_0, T - t_0)$ by $L_v(x_0, T - t_0)$. A solution of $\Gamma_v(x_0, T - t_0)$ will be for us a subset $W_v(x_0, T - t_0) \subset L_v(x_0, T - t_0)$ of imputations which satisfies additional "optimality" conditions.

An allocation $\xi = (\xi_1, \ldots, \xi_n)$ represents an equitable imputation if each player receives at least a maximal guaranteed payoff and the entire maximal payoff is distributed evenly without a remainder.

1.4.2 The Principle of Time-Consistency (Dynamic Stability)

The formalization of the notion of optimal behaviour constitutes one of the fundamental problems in the theory of *n*-person games. At present, for various classes of games different solution concepts are constructed. The behaviour of the players in the game is characterised by their strategies, if the game is non-cooperative, or by their imputations, if the game is cooperative. Behaviour that satisfies a given optimality principle is called a solution of the game in the sense of this principle and must possess two properties. First, it must be feasible in the game; second, it must adequately reflect the conceptual notion of optimality, taking into account the special features of the class of games for which it is defined.

For dynamic games, it is natural to add one more requirement: feasibility and purposefulness of the optimality principle are to be preserved throughout the game. This requirement is called time-consistency or dynamic stability of a solution of the game (see Yeung and Petrosyan 2006; Haurie 1976; Petrosyan 1977; Petrosyan and Danilov 1979; Petrosyan and Zaccour 2003).

Time-consistency of a solution of a differential game means the following: if the game proceeds along a "conditionally optimal" trajectory, and players are guided by the same optimality principle at each instant of time, they never have an incentive to deviate from the previously adopted "optimal" behaviour. Time-consistency is violated if at some point in time continuation of the initial behaviour becomes non-optimal and hence the initially chosen solution proves to be unfeasible.

Assume now that at the start of the game the players have adopted an optimality principle and have constructed a solution based on it; that is, they have chosen an imputation satisfying the chosen principle of optimality, say the core, nucleolus, Neumann–Morgenstern solution etc. From the definition of cooperative game it follows that the evolution of the game is to be along the trajectory providing a maximal total payoff for the players. When moving along this "conditionally optimal" trajectory, the players pass through subgames with current initial states and current duration. In due course, not only the conditions of the game and the opportunities of the players, but even the players' interests may change: at an instant t the optimal solution of the current game may not keep to the initially chosen "conditionally optimal" trajectory. If this occurs, the chosen optimality principle is time-inconsistent and, as a result, the motion itself is dynamically unstable; we have seen the time-inconsistency of the Shapley value in the last example of the previous section.

From now on, we focus our attention on time-consistent solutions of cooperative differential games.

Choose therefore an optimality principle in the game $\Gamma_v(x_0, T - t_0)$. Denote by $W_v(x_0, T - t_0)$ the solution of this game, based on the chosen principle of optimality and constructed in the initial state $x(t_0) = x_0$. The set $W_v(x_0, T - t_0)$, which is assumed to be nonempty, is a subset of the imputation set $L_v(x_0, T - t_0)$. Denote by $x^*(t)$, for $t \in [t_0, T]$, the conditionally optimal trajectory; by definition of conditional optimality, players obtain the largest total payoff along this trajectory. We assume henceforth that such a trajectory exists.

Consider now the behaviour of the set $W_v(x_0, T - t_0)$ along the conditionally optimal trajectory. For each current state $x^*(t)$, we define the characteristic function $v(S; x^*(t), T - t)$ as the value of the zero-sum differential game $\Gamma_S(x^*(t), T - t)$ between coalitions S and $N \backslash S$ from the initial state $x^*(t)$ and with duration $T - t$, as it was done already for the game $\Gamma(x_0, T - t_0)$ above.

The current cooperative subgame $\Gamma_v(x^*(t), T - t)$ is, as before, defined as the pair $\big(N, v(S, x^*(t), T - t)\big)$. The imputation set in the game $\Gamma_v(x^*(t), T - t)$ is of the form:

$$L_v(x^*(t), T - t) = \Bigg\{ \xi \in R^n \ : \ \xi_i \geq v(\{i\}; x^*(t), T - t), \ i = 1, \ldots, n; \\ \sum_{i \in N} \xi_i = v(N; x^*(t), T - t) \Bigg\},$$

where

$$v(N; x^*(t), T - t) = v(N; x_0, T - t_0) - \int_{t_0}^{t} \sum_{i \in N} h_i(x^*(\tau)) d\tau.$$

In this expression, the quantity

$$\int_{t_0}^{t} \sum_{i \in N} h_i(x^*(\tau)) d\tau$$

can be interpreted as the total gain of the players on the time interval $[t_0, t]$ when the motion is carried out along the trajectory $x^*(t)$.

We have in this way obtained a family of current games

$$\Gamma_v(x^*(t), T - t) = \big(N, v(S; x^*(t), T - t)\big), \quad \text{parametrised by} \quad t \in [t_0, T],$$

which is parametrised by $t \in [t_0, T]$ and which is defined along the conditionally optimal trajectory $x^*(t)$; the corresponding family of solutions $W_v(x^*(t), T - t)$ has the property that each solution is generated by the same principle of optimality as the initial solution $W_v(x_0, T - t_0)$.

It is obvious that the set $W_v(x^*(T), 0)$ is a solution of terminal game $\Gamma_v(x^*(T), 0)$ and that it is composed of the only possible imputation

$$H(x^*(T)) = \big(H_1(x^*(T)), \dots, H_n(x^*(T)\big),$$

where $H_i(x^*(T))$ is the terminal payoff of player i along the trajectory $x^*(t)$.

1.4.3 Time-Consistent Solutions

Let the conditionally optimal trajectory $x^*(t)$ be such that

$$W_v(x^*(t), T - t) \neq \emptyset, \quad t_0 \leq t \leq T.$$

If this condition is not satisfied, it is impossible for players to adhere to the chosen optimality principle, since at the first instant t for which the condition is violated, the players have no possibility to follow the principle any longer.

Assume that in the initial state x_0 the players agree upon the imputation $\xi^0 \in W_v(x_0, T - t_0)$. This means that in the state x_0 the players agree upon such an allocation of the total maximal gain that, when the game terminates at the instant T, the share of ith player is equal to ξ_i^0, i.e. the ith component of the imputation ξ^0. Suppose player i's payoff—his share—after the time interval $[t_0, t]$ has elapsed is equal to $\xi_i(x^*(t))$. Then on the remaining time interval $[t, T]$ he has to receive the gain $\eta_i^t = \xi_i^0 - \xi_i(x^*(t))$ in order to be consistent with ξ^0. For the original agreement—the imputation ξ^0—to remain in force at the instant t, it is therefore necessary that

$$\eta^t = (\eta_i^t, \dots, \eta_n^t) \in W_v(x^*(t), T - t),$$

i.e. that η^t is a solution of the current subgame $\Gamma_v(x^*(t), T - t)$. If such a condition can be satisfied for each $t \in [t_0, T]$ along the trajectory $x^*(t)$, then the imputation ξ^0 can be realized. This is the conceptual idea of time-consistent imputations.

Restricted to the time interval $[t, T]$, $t_0 \leq t \leq T$, the coalition N obtains the payoff

$$v(N; x^*(t), T - t) = \sum_{i \in N} \left[\int_t^T h_i(x^*(\tau))d\tau + H_i(x^*(T)) \right]$$

along the trajectory $x^*(t)$. The difference

$$v(N; x_0, T - t_0) - v(N; x^*(t), T - t) = \int_{t_0}^t \sum_{i \in N} h_i(x^*(\tau))d\tau$$

is then equal to the payoff the coalition N obtains over the time interval $[t_0, t]$. The share of the ith player in this payoff, assuming that payoffs are transferable, can be represented as

$$\gamma_i(t) = \int_{t_0}^{t} \beta_i(\tau) \sum_{j=1}^{n} h_j(x^*(\tau))d\tau = \gamma_i(x^*(t), \beta), \tag{1.12}$$

where the function $\beta_i(\tau)$ is integrable over $[t_0, T]$ and satisfies the condition

$$\sum_{i=1}^{n} \beta_i(\tau) = 1, \quad t_0 \leq \tau \leq T. \tag{1.13}$$

From (1.12) we immediately obtain

$$\frac{d\gamma_i}{dt} = \beta_i(t) \sum_{j \in N} h_j(x^*(t)).$$

This quantity may be interpreted as the instantaneous gain of player i at the moment t. Hence it is clear that the vector $\beta(t) = (\beta_1(t), \ldots, \beta_n(t))$ prescribes the distribution of the total gain among the members of coalition N. By properly choosing $\beta(t)$, the players can ensure the desirable outcome, i.e. they can regulate how the players receive their gains over time. They can do this in such a way that at no instant $t \in [t_0, T]$ there will be an objection against the realization of the original agreement, that is, against the imputation ξ^0.

Definition 9. The imputation $\xi^0 \in W_v(x_0, T - t_0)$ is called time-consistent in the game $\Gamma_v(x_0, T - t_0)$ if the following conditions are satisfied:

(1) there exists a conditionally optimal trajectory $x^*(t)$ along which

$$W_v(x^*(t), T - t) \neq \emptyset$$

for $t_0 \leq t \leq T$,

(2) there exists a vector function $\beta(t) = (\beta_1(t), \ldots, \beta_n(t))$, integrable over $[t_0, T]$, such that for each $t_0 \leq t \leq T$, $\sum_{i=1}^{n} \beta_i(t) = 1$ and

$$\xi^0 \in \bigcap_{t_0 \leq t \leq T} [\gamma(x^*(t), \beta) \oplus W_v(x^*(t), T - t)], \tag{1.14}$$

where $\gamma(x^*(t), \beta) = (\gamma_1(x^*(t), \beta), \ldots, \gamma_n(x^*(t), \beta))$.

The cooperative differential game $\Gamma_v(x_0, T - t_0)$ with side payments has a time-consistent solution $W_v(x_0, T - t_0)$ if all of the imputations $\xi \in W_v(x_0, T - t_0)$ are time-consistent.

The conditionally optimal trajectory along which there exists a time-consistent solution of the game $\Gamma_v(x_0, T - t_0)$ is called an optimal trajectory.

For $t = T$, time-consistency implies that $\xi_0 \in \gamma(x^*(T), \beta) \oplus W_v(x^*(T), 0)$. Recall that $W_v(x^*(T), 0)$ is a solution of the current game $\Gamma_v(x^*(T), 0)$; this game occurs at the last moment $t = T$ of the trajectory $x^*(t)$, has a duration 0 and is made up of the only imputation $\xi^T = H(x^*(T)) = (H_1(x^*(T)), \dots, H_n(x^*(T)))$. The imputation ξ^0 may therefore be represented as

$$\xi^0 = \gamma(x^*(T), \beta) + H(x^*(T)) \tag{1.15}$$

or

$$\xi^0 = \int_{t_0}^{T} \beta(\tau) \sum_{i \in N} h_i(x^*(\tau)) d\tau + H(x^*(T)).$$

The time-consistent imputation $\xi^0 \in W_v(x_0, T - t_0)$ may be realized as follows. From (1.14) at any instant $t_0 \leq t \leq T$ we have

$$\xi^0 \in \gamma(x^*(t), \beta) \oplus W_v(x^*(t), T - t), \tag{1.16}$$

This relation implies the existence of a vector $\xi^t \in W_v(x^*(t), T - t)$ such that $\xi^0 = \gamma(x^*(t), \beta) + \xi^t$. By combining (1.12) and (1.15), we see that

$$\xi^t = \xi^0 - \gamma(x^*(t), \beta) = \int_{t}^{T} \beta(\tau) \sum_{i \in N} h_i(x^*(\tau)) d\tau + H(x^*(T)).$$

The integrand in this expression, denoted

$$\alpha_i(\tau) = \beta_i(\tau) \sum_{j \in N} h_j(x^*(\tau)), \quad i \in N,$$

is called the imputation distribution procedure (IDP).

We conclude that for any vector valued function $\beta(t)$ satisfying conditions (1.12) and (1.13) at each time instant $t_0 \leq t \leq T$, the players are guided by the imputation $\xi^t \in W_v(x^*(t), T - t)$ and the associated optimality principle throughout the game.

Let us make the following additional assumption.

Assumption A. *The vector $\xi^t \in W_v(x^*(t), T - t)$ may be chosen as a continuously differentiable function of the argument t.*

We shall show that under assumption A, we can always ensure time-consistency of the imputation $\xi^0 \in W_v(x_0, T - t_0)$ by properly choosing $\beta(t)$.

To see this, choose $\xi^t \in W_v(x^*(t), T - t)$ such that it is a continuously differentiable function of t, which is possible according to the assumption. Constructing the difference $\gamma(t) = \xi^0 - \xi^t$, we obtain that

$$\xi^t + \gamma(t) \in W_v(x_0, T - t_0).$$

We are looking for a vector function $\beta(t) = (\beta_1(t), \ldots, \beta_n(t))$, integrable on $[t_0, T]$, that satisfies conditions (1.12), (1.13). For the sake of simplicity, we shall write $\gamma(t)$ instead of $\gamma(x^*(t), \beta)$. Rewriting (1.12) in vector form yields

$$\int_{t_0}^{t} \beta(\tau) \sum_{i \in N} h_i(x^*(\tau) d\tau = \gamma(t).$$

By differentiating with respect to t and rearranging terms we obtain the following expression for $\beta(t)$

$$\beta(t) = \frac{1}{\sum_{i \in N} h_i(x^*(t))} \cdot \frac{d\gamma(t)}{dt} = -\frac{1}{\sum_{i \in N} h_i(x^*(t))} \cdot \frac{d\xi^t}{dt}; \tag{1.17}$$

the last equality follows from differentiating the identity

$$\xi^0 = \gamma(t) + \xi^t.$$

Taking (1.17) as the definition of $\beta(t)$, it is clear that this function is continuous. It remains to check that condition (1.13) is satisfied. Indeed, since

$$\sum_{i \in N} \xi_i^t = v(N; x^*(t), T - t),$$

it follows that

$$\sum_{i \in N} \beta_i(t) = -\frac{\sum_{i \in N} \frac{d\xi_i^t}{dt}}{\sum_{i \in N} h_i(x^*(t))} = -\frac{\frac{d}{dt} v(N; x^*(t), T - t)}{\sum_{i \in N} h_i(x^*(t))}$$

$$= -\frac{\frac{d}{dt}\left[\sum_{i \in N} \left(\int_t^T h_i(x^*(\tau)) d\tau + H_i(x^*(T)) \right) \right]}{\sum_{i \in N} h_i(x^*(t))} = \frac{\sum_{i \in N} h_i(x^*(t))}{\sum_{i \in N} h_i(x^*(t))} = 1.$$

We have proved the following theorem.

Theorem 2. *If assumption A is satisfied and for all $t \in [t_0, T]$ the condition*

$$W_v(x^*(t), T - t) \neq \emptyset \tag{1.18}$$

is satisfied, then the solution $W_v(x_0, T - t_0)$ is time-consistent.

The main problem in subsequent theoretical developments is to find conditions that can be imposed on the vector function $\beta(t)$, which ensure time-consistency of specific solution forms $W_v(x_0, T - t_0)$ in various classes of games.

In the remainder of this section, we discuss the new concept of strong time-consistency, and we define time-consistent solutions for cooperative games with terminal payoffs.

1.4.4 Strongly Time-Consistent Solutions

Recall that for a time-consistent imputation $\xi^0 \in W_v(x_0, T - t_0)$, by definition there are an integrable vector valued function $\beta(t)$ and an imputation ξ^t from the solution $W_v(x^*(t), T - t)$ of the current game $\Gamma_v(x^*(t), T - t)$, neither of which is in general unique, such that

$$\xi^0 = \gamma(x^*(t), \beta) + \xi^t$$

for each $t \in [t_0, T]$, where $\gamma(x^*(t), \beta)$ is the vector of total payoffs to the players up to time t.

The condition of time-consistency does not affect imputations from the set $W_v(x^*(t), T - t)$ that fail to satisfy this equation. It is now interesting to consider the situation where any imputation from the current solution $W_v(x^*(t), T - t)$ provides a "good" continuation of the original agreement. This situation obtains if $\xi^0 \in W_v(x_0, T - t_0)$ is a time-consistent imputation and if for every $\xi^t \in W_v(x^*(t), T - t)$, the condition

$$\gamma(x^*(t), \beta) + \xi^t \in W_v(x_0, T - t_0),$$

is satisfied.

By slightly strengthening this requirement, we obtain the concept of strong time-consistency.

Definition 10. An imputation $\xi^0 \in W_v(x_0, T - t_0)$ is called strongly time-consistent in the game $\Gamma_v(x_0, T - t_0)$, if the following conditions are satisfied:

(1) the imputation ξ^0 is time-consistent;
(2) for any $t_0 \le t_1 \le t_2 \le T$ and for $\beta(t)$ corresponding to the imputation ξ^0 we have

$$\gamma(x^*(t_2), \beta) \oplus W_v(x^*(t_2), T - t_2) \subset \gamma(x^*(t_1), \beta) \oplus W_v(x^*(t_1), T - t_1). \quad (1.19)$$

A cooperative differential game $\Gamma_v(x_0, T - t_0)$ with side payments has a strongly time-consistent solution $W_v(x_0, T - t_0)$ if all imputations from this solution are strongly time-consistent.

1.4.5 Terminal Payoffs

In this section, we consider the situation that the players do only obtain payoffs at the termination of the game, that is

$$K_i(x_0, T - t_0; u_1, \dots, u_n) = H_i(x(T)), \qquad i = 1, \dots, n;$$

this expression is obtained from (1.11) by setting $h_i \equiv 0$ for all i. The resulting cooperative differential game with the terminal payoffs is again denoted by $\Gamma_v(x_0, T - t_0)$. In the following, we write

$$H(x^*(T)) = \big(H_1(x^*(T)), \dots, H_n(x^*(T))\big),$$

for the vector whose component are the payoffs at the end point of the conditionally optimal trajectory.

Theorem 3. *In the cooperative differential game $\Gamma_v(x_0, T - t_0)$ with terminal payoffs $H_i(x(T))$, the vector $\xi^0 = H(x^*(T))$ is the unique time-consistent imputation.*

Proof. Time-consistency of the imputation $\xi^0 \in W_v(x_0, T - t_0)$ implies that

$$\xi^0 \in \bigcap_{t_0 \le t \le T} W_v(x^*(t), T - t).$$

But since the current game $\Gamma_v(x^*(T), 0)$ is of zero duration,

$$L_v(x^*(T), 0) = W_v(x^*(T), 0) = H(x^*(T)).$$

Hence

$$\bigcap_{t_0 \le t \le T} W_v(x^*(t), T - t) = H(x^*(T)),$$

i.e. $\xi^0 = H(x^*(T))$ and there are no other imputations.

Theorem 4. *For a time-consistent solution to exist in the game with terminal payoffs it is necessary and sufficient that for all $t_0 \le t \le T$*

$$H(x^*(T)) \in W_v(x^*(t), T - t).$$

This theorem is a corollary of the previous one.

Thus, if in the game with terminal payoffs there is a time-consistent imputation, then in the initial state x_0 the players have to agree upon the imputation $H(x^*(T)) \in W_v(x_0, T - t_0)$; moreover, while moving along the optimal trajectory $x^*(t)$, for each $t_0 \le t \le T$ this imputation belongs to the solution of the current game $\Gamma_v(x^*(t), T - t)$.

The theorem states that in the game with terminal payoffs only a single imputation from the set $W_v(x_0, T - t_0)$ can be time-consistent. This is a highly improbable event, as this means that imputation $H(x^*(T))$ belongs to the solution of every subgame along the conditionally optimal trajectory. For such games the notions of time-consistency and strong time-consistency of the solution $W_v(x_0, T - t_0)$ is therefore largely irrelevant.

It is important to mention that a theorem analogous to Theorem 1 is true for an n-person differential game (see Petrosyan and Zenkevich 2009). This shows that dynamic games are a very effective tool for constructing the bridge between noncooperative and cooperative game theory.

Acknowledgements This work was supported by the St. Petersburg State University under research grant No. 9.38.245.2014. The author thanks Florian Wagener for many valuable comments.

References

Basar T, Olsder GJ (1982) Dynamic noncooperative game theory. Academic, London

Berkovitz LD (1964) A variational approach to differential games. In: Dresher M, Shapley LS, Tucker AW (eds) Advances in game theory. Annals of mathematics studies, vol 52. Princeton University Press, Princeton, pp 127–174

Dutkevich YG, Petrosyan LA (1972) Games with a “life-line”. The case of l-capture. SIAM J Control 10(1):40–47

Fleming WH (1961) The convergence problem for differential games. J Math Anal Appl 3:102–116

Haurie A (1976) A note on nonzero-sum differential games with bargaining solutions. J Optim Theory Appl 18:31–39

Isaacs R (1965) Differential games. Wiley, New York

Krasovskii NN (1963) On a problem of tracking. Prikl Mat i Mekh 27(2):244–254 (Russian)

Kuhn HW (1953) Extensive games and the problem of information. In: Kuhn HW, Tucker AW (eds) Contributions to the theory of games II. Annals of mathematics studies, vol 28. Princeton University Press, Princeton, pp 193–216

Petrosyan LA (1965) A family of differential survival games in the space R^n. Dokl Akad Nauk SSSR 161(1):52–54 (Russian)

Petrosyan LA (1977) Stability of solutions of nonzero sum differential games with many participants. Vestn Leningr Univ 19:6–52 (Russian)

Petrosyan LA, Danilov, NN (1979) Stability of solutions in non-zero sum differential games with transferable payoffs. Vestn Leningr Univ 1:52–59 (Russian)

Petrosyan LA, Murzov NV (1967) Tag differential game with many participants. Vestn Leningr Univ 13:125–129 (Russian)

Petrosyan LA, Shiryaev VD (1978) Simple pursuit by one pursuer of two pursued. In: Some questions of differential and integral equations and their applications, vol 3. Yakutsk Gos Univ, Yakutsk, pp 103–108 (Russian)

Petrosyan LA, Zaccour G (2003) Time-consistent Shapley value allocations of pollution cost reduction. J Econ Dyn Control 27(3):381–398

Petrosyan LA, Zenkevich NA (2009) Principles of stable cooperation. Mat Teor Igr Prilozh 1(1):102–117 (Russian)

Pontryagin LS (1967) Linear differential games I. Dokl Akad Nauk SSSR 174(6):1278–1280 (Russian)

Shapley LS (1953) A value for *n*-person games. In: Kuhn HW, Tucker AW (eds) Contributions to the theory of games II. Annals of mathematics studies, vol 28. Princeton University Press, Princeton, pp 307–317

Steinhaus H (1925) Definitions of a theory of games and pursuit. Mysl Akademicka, 1(1). Reprinted in: Steinhaus H (1960) Naval Res Logist Quart 7:105–107

Yeung DWK, Petrosyan LA (2006) Cooperative stochastic differential games. Springer, New York

Chapter 2
Dynamic Admission Game into an M/M/1 Queue

Eitan Altman and Tania Jiménez

Abstract Around 50 years ago, P. Naor has derived the optimal social and individual admission rules to an M/M/1 queue. In both cases, the optimal policies were identified to be of a pure threshold type: admit if and only if the number queued upon arrival is below some threshold. The value of the threshold in the individual optimal case was shown to be larger than the one for the social optimal criterion. We make the observation that admitting according to a threshold policy requires only the information of whether the queue is above or below a threshold. We call these "red" and "green" light, respectively, associated with a threshold, say L. The question that we pose in this paper is: what happens if one restricts to the above information pattern but let the threshold level L be chosen by the system which signals to arrivals whether the queue is above or below the threshold. Can one find a choice of a threshold that will induce an equilibrium that performs better than in the case that full information is available? We also examine the question of what is the threshold that maximizes the revenue for the queue. We show that the choice of threshold that maximizes the system's performance at equilibrium is the same as under the full information case if the service in the queue follows the FIFO discipline.

Keywords M/M/1 • Game Theory • Threshold policies

E. Altman
INRIA, 2004 Route des Lucioles, 06902 Sophia Antipolis, France

LINCS - Laboratory of Information, Network and Communication Sciences Paris
e-mail: eitan.altman@inria.fr

T. Jiménez (✉)
Avignon University, LIA EA4128, F-84000 Avignon, France
e-mail: tania.jimenez@univ-avignon.fr

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*,
Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_2

2.1 Introduction

This paper is devoted to revisiting the problem of whether an arrival should queue or not in an M/M/1 queue. To the best of our knowledge, this is the first problem to be studied in optimal control of queues, going back to the seminal paper of Naor (1969). Naor considered an M/M/1 queue, in which a controller has to decide whether arrivals should enter a queue or not. The objective was to minimize a cost with the form of a weighted difference between the average expected waiting time of those that enter, and the acceptance rate of customers. The strategy that minimizes the above cost was shown to be of a threshold type where arrivals are accepted as long as the queue size does not exceed some threshold L and are otherwise denied access.

Naor then considered the individual optimality problem in which an arrival at the queue decides whether to enter the queue and receive service. The individual knowing the size i of the queue, joins the queue if its expected waiting time does not exceed the value of being served. The solution to this individual optimization problem can be viewed as a Nash equilibrium in a non-cooperative game between the players. A simple argument shows that this equilibrium is also of a threshold type with a threshold L' which satisfies $L' \geq L$. Thus, under individual optimality, arrivals that join the queue wait longer in average.

Finally, he showed that there exists some toll such that if it is imposed on arrivals for joining the queue then the threshold value of the individually optimal policy can be made to agree with the social optimal one. Since this seminal work of Naor there has been a huge amount of research that extend the model. More general interarrival and service times have been considered, more general networks, other objective functions and other queuing disciplines, see e.g. Yechiali (1971), Stidham and Weber (1989), Stidham et al. (1995), Hsiao and Lazar (1991), Korilis and Lazar (1995), Altman and Shimkin (1998), Hordijk and Spieksma (1989), Altman et al. (2000), Stidham (1985) and references therein.

In the original work of Naor, the decision maker(s) have full state information when entering the system. However, the fact that a threshold policy is optimal implies that for optimally controlling arrivals we only need partial information—we need a signal to indicate whether the queue exceeds or not the threshold value L. The fact that this much simpler information structure is sufficient for obtaining the same performance as in the full information case motivates us to study the performance of threshold policy and related optimization issues.

We first consider the socially optimal control policy for a given (non-necessarily optimal) threshold value L. When L is chosen non-optimally then the optimal policy for the partial information problem does not anymore coincide with the policy with full information.

We then study the individual optimization problem with the same partial information: a signal (red) if the queue length exceeds some value L and a green signal otherwise.

For both the social and the individual optimization problems we show that the following structure holds: either whenever the signal is green all arrivals are accepted with probability 1, or whenever the signal is red all arrivals are rejected with probability 1.

We note that by using this signaling approach instead of providing full state information, users cannot choose any threshold policy with parameter different than L. Thus, in the individual optimization case, one could hope that by determining the signaling according to the value L that optimizes the socially optimal problem (in case of full information), one would obtain the socially optimal performance. We show that this is not the case, and determine the value L for which the reaction of the users optimizes the system performance. We compare this to the performance in case of full information.

2.2 The Model

2.2.1 The General Threshold Framework

Consider an M/M/1/k queue, i.e. a single server queue with an independent Poisson arrival process, independent exponentially distributed service times provided according to the FCFS (first come first served) regime, and with a storage capacity of k customers; both finite and infinite k will be considered. The input rate of admitted customer is state dependent. We are in particular interested in threshold admission rates given by $\underline{\lambda}$ for $i \geq L$ and by $\overline{\lambda}$ otherwise. Here i, the number of queued customer at the queue, is taken to be the state of the queue. With this as the state, the queue process is Markovian. If we take $\underline{\lambda} = 0$ then we obtain the model of Naor. $\underline{\lambda}$ can be interpreted as some non-controlled (or non strategic) flow which is always accepted at the queue. The rest of the flow (of rate $\overline{\lambda} - \underline{\lambda}$) is fully controlled.

Let μ be the service rate and set $\overline{\rho} = \overline{\lambda}/\mu$ and $\underline{\rho} = \underline{\lambda}/\mu$. For the case of infinite k we shall make the standard stability assumption that $\underline{\rho} < 1$. Without loss of generality we may assume that $L \leq k$.

The Optimal Control Framework We assume that $\nu < \underline{\lambda}$ is the rate of some uncontrolled Poisson flow. In addition there is an independent Poisson arrival flow of intensity $\zeta \geq 0$. We restrict to stationary policies, i.e. policies that are only function of the observation. A policy is thus a set of two probabilities: q_s where s is either R or G. q_s is the probability of accepting an arrival when the signal is s. Define $\mathbf{q} = (q_G, q_R)$.

$$\underline{\lambda} = \nu + \zeta q_R, \quad \overline{\lambda} = \nu + \zeta q_G. \tag{2.1}$$

The Non-cooperative Game Problem We again assume that there is some uncontrolled flow ν and a flow of identical strategic players with intensity ζ. All users receive the signal G or R—as before—and we restrict to stationary polices as in the optimal control case.

Assume that an arrival has a reward $\gamma > 0$ for being processed in the queue, and a waiting cost of $E_{\mathbf{q}}[W|s]$ where W is the waiting time, and s corresponds to the signal that the customer receives upon arrival. Note that $E_{\mathbf{q}}[W|s] = E_{\mathbf{q}}[I|s]/\mu$ where I stands for the number of customers at the queue ahead of the arrival.

Let $Y(\mathbf{q})$ be the set of best responses of an individual if all the rest use $\mathbf{q}$ and the system is in the corresponding steady state.

Then $\mathbf{q}$ is an equilibrium strategy if and only if $\mathbf{q} \in Y(\mathbf{q})$.

2.3 Steady State Probabilities for Threshold Policies

We compute the steady state probabilities as well as the performance measures that will appear in the utilities within the general framework. They will be used in both the optimal control framework as well as in the game framework.

The balance equations are given

$$\mu\pi(i+1, L) = \lambda(i)\pi(i, L)$$

where $\lambda(i) = \underline{\lambda}$ for $i \geq L$ and is otherwise given by $\lambda(i) = \overline{\lambda}$. The solution of these equations give

$$\pi(i, L) = \pi(0, L)\overline{\rho}^i$$

for $i \leq L$ and otherwise

$$\pi(i, L) = \pi(L, L)\underline{\rho}^{i-L}. \tag{2.2}$$

Hence

$$\begin{aligned} \pi(0, L) &= \frac{1}{\sum_{i=0}^{L-1} \overline{\rho}^i + \overline{\rho}^L \sum_{i=0}^{k-L} \underline{\rho}^i} \\ &= \frac{1}{\frac{1-\overline{\rho}^L}{1-\overline{\rho}} + \frac{\overline{\rho}^L}{1-\underline{\rho}}(1 - \underline{\rho}^{k-L+1})} \end{aligned}$$

Thus

$$\pi(L, L) = \pi(0, L) \cdot \overline{\rho}^L = \frac{\overline{\rho}^L}{\frac{1-\overline{\rho}^L}{1-\overline{\rho}} + \frac{\overline{\rho}^L}{1-\underline{\rho}}(1 - \underline{\rho}^{k-L+1})} \tag{2.3}$$

Assume that an arrival receives the information on whether the size of the queue exceeds $L - 1$ or not. If it does not exceeds we shall say that it receives a "green" signal denoted by G, and otherwise a red one (R). The conditional state probabilities given the signals are denoted by

$$\pi(i, L|R) = \underline{\rho}^{i-L} \frac{1-\underline{\rho}}{1-\underline{\rho}^{k-L+1}}$$

for $k \geq i \geq L$, and is otherwise zero. The conditional tail distribution is

$$P(I > n|R) = \underline{\rho}^{n+1-L} \frac{1-\underline{\rho}^{k-n+1}}{1-\underline{\rho}^{k-L+1}}$$

for $n \geq L-1$, and is otherwise 1. Thus

$$E(I|R) = L - 1 + \frac{1}{1-\underline{\rho}} \tag{2.4}$$

when $k = \infty$ and

$$E(I|R) = L - 1 + \frac{1}{1-\underline{\rho}} - \frac{(k-L)\cdot\underline{\rho}^{k-L+1}}{1-\underline{\rho}^{k-L+1}} \tag{2.5}$$

otherwise.

For a green light we have:

$$\pi(i, L|G) = \frac{1-\overline{\rho}}{1-\overline{\rho}^L}\overline{\rho}^i$$

for $0 \leq i < L$ and is otherwise zero. Hence the tail probabilities are

$$P(I > n|G) = \frac{\overline{\rho}^{n+1} - \overline{\rho}^L}{1-\overline{\rho}^L}$$

for $n < L$, and is otherwise 0. Hence

$$E(I|G) = \sum_{n=0}^{L-1} P(I > n|G) = \frac{1}{1-\overline{\rho}^L}\left(\frac{(\overline{\rho}^L - \overline{\rho})}{\overline{\rho}-1} - (L-1)\overline{\rho}^L\right) \tag{2.6}$$

In the sequel we shall add often $\mathbf{q}$ in order to stress the dependence of various quantities on $\mathbf{q}$. The dependence is through $\overline{\rho}$ and $\underline{\rho}$ [due to (2.1)].

2.4 The Partially Observed Control Problem

The global optimisation problem is to minimize $J_{\mathbf{q}}$ over $\mathbf{q}$ where $J_{\mathbf{q}}$ is the following weighted difference between the expected queueing cost and a payoff per rate of accepted customers:

$$J_{\mathbf{q}} = E_{\mathbf{q}}[I] - \gamma T_{acc}(\mathbf{q}) = \sum_{s=G,R} P_{\mathbf{q}}(s)(E_{\mathbf{q}}[I|s] - \gamma T_{acc}(\mathbf{q})) \tag{2.7}$$

where

$$T_{acc} = \overline{\lambda} \cdot P(G) + \underline{\lambda} \cdot P(R) = \mu[P(R)(\underline{\rho} - \overline{\rho}) + \overline{\rho}]$$

and $P(R) = P(I \geq L)$ is given by

$$P(R) = \pi(L, L) \frac{1 - \underline{\rho}^{k-L+1}}{1 - \underline{\rho}} \tag{2.8}$$

$$\begin{aligned}
E[I] &= E[I|R] \cdot P(R) + E[I|G] \cdot P(G) = (E[I|R] - E[I|G]) \cdot P(R) + E[I|G],\\
&= \Big(L - 1 + \frac{1}{1-\underline{\rho}} - \frac{(k-L) \cdot \underline{\rho}^{k-L+1}}{1 - \underline{\rho}^{k-L+1}} - \frac{1}{1-\overline{\rho}^L}\left(\frac{(\overline{\rho}^L - \overline{\rho})}{\overline{\rho} - 1} - (L-1)\overline{\rho}^L\right)\Big)\\
&\times \frac{\overline{\rho}^L}{\frac{1-\overline{\rho}^L}{1-\overline{\rho}} + \frac{\overline{\rho}^L}{1-\underline{\rho}}(1 - \underline{\rho}^{k-L+1})} \left(\frac{1 - \underline{\rho}^{k-L+1}}{1 - \underline{\rho}}\right) + \frac{1}{1-\overline{\rho}^L}\left(\frac{(\overline{\rho}^L - \overline{\rho})}{\overline{\rho} - 1} - (L-1)\overline{\rho}^L\right)
\end{aligned}$$

when $k \neq \infty$ and

$$\begin{aligned}
&= (((L-1) + \frac{1}{\underline{\rho}^L(1-\underline{\rho})}) - (\frac{1}{1-\overline{\rho}^L}\left(\frac{(\overline{\rho}^L - \overline{\rho})}{\overline{\rho} - 1} - (L-1)\overline{\rho}^L\right)))\\
&\times \frac{(1-\overline{\rho})\overline{\rho}^L}{(1-\underline{\rho}) + \overline{\rho}^L(\underline{\rho} - \overline{\rho})} + \Big(\frac{1}{1-\overline{\rho}^L}\left(\frac{(\overline{\rho}^L - \overline{\rho})}{\overline{\rho} - 1} - (L-1)\overline{\rho}^L\right)\Big)
\end{aligned}$$

when $k = \infty$.

The expression obtained for $J_{\mathbf{q}}$ is lower semi-continuous in the policy $\mathbf{q} = \{q_s, s = G, R\}$. Hence a minimizing policy $\mathbf{q}^*$ exists.

Lemma 1. *Consider the case of $k = \infty$. Assume that $\nu > 0$. If $\underline{\rho}_R \geq 1$ then for any L, $E[I]$ is infinite.*

Proof. The expected queue length $E[I_t]$ at any time t and for any L is bounded from below by $E[I'_t] - L$ where I'_t is the queue size obtained when replacing q_G with $q_G = q_R$. $E[I'_t]$ corresponds to an M/M/1 queue with a workload $\rho \geq 1$ which is known to have infinite expectation. ∎

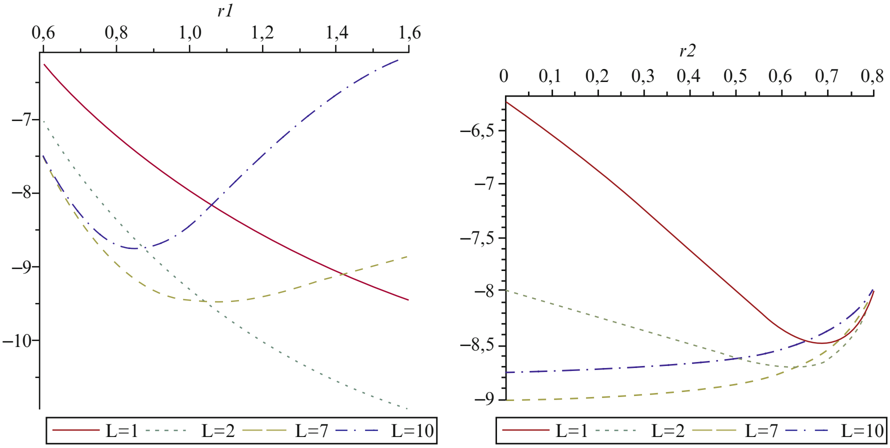

Fig. 2.1 The performance (J_q in y-axis and $\underline{\rho}$ and $\overline{\rho}$ x-axis) of different policies for several values of L for $k = \infty$

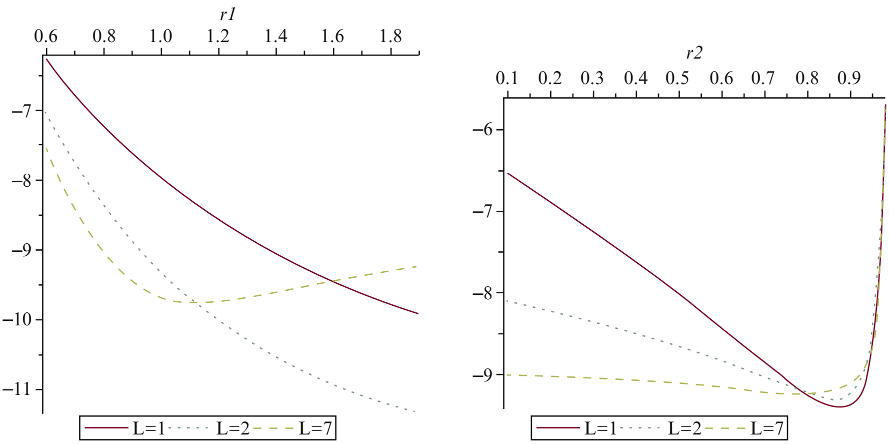

Fig. 2.2 The performance (J_q in y-axis and $\underline{\rho}$ and $\overline{\rho}$ x-axis) of different policies for several values of L for $k = 8$

2.4.1 The Structure of Optimal Policies

Figure 2.1 shows the values of $J_{\mathbf{q}}$ for various values of $\underline{\rho}$ and $\overline{\rho}$ for k infinite. We assume that ν and λ are such that $\overline{\rho} = 0.8$ and $\underline{\rho} = 0.3$. We further took $\mu = 1$, $\gamma = 15$, for four different values of the threshold L. In the left part of the figure, $\overline{\rho}$ is varied while keeping $\underline{\rho} = 0.3$ constant. In the right part of the figure $\underline{\rho}$ is varied while keeping $\overline{\rho} = 0.8$ constant. Similar plots are given in Fig. 2.2 for the case of finite storage capacity of $k = 8$.

Consider the case of $L = 10$ in Fig. 2.1. Consider the case of $\underline{\rho} = 0.3, \overline{\rho} = 0.8$. At these fixed values, it is seen that no unilateral change in either $\underline{\rho}$ or in $\overline{\rho}$ can further increase $J_{\mathbf{q}}$. This is thus a (locally) optimal policy. It is seen to satisfy the following: q_R is an interior point while $q_L = 0$ is on the boundary.

We shall show the following more general structure. For any L the optimal vector $\mathbf{q}$ satisfies the following property: whenever the minimum cost is achieved at an interior point for one of the components of $\mathbf{q}$, then it is achieved on the boundary for the other component. We shall next prove this structure for the partially observable control problem with $k = \infty$.

Theorem 1. *Consider $k = \infty$. Assume that $0 < \nu/\mu < 1$. Then there is a unique optimal stationary strategy and it has the following property: either $q^*(G) = 1$ or $q^*(R) = 0$.*

Proof. Let $\mathbf{q}$ be optimal. We first show that $\alpha > 0$ where $\alpha := \mu - (\nu + q_R\zeta)$. Indeed, if it were not the case then we would have $\rho \geq 1$ so by the previous lemma, the queue length and hence the cost would be infinite. But then $\mathbf{q}$ cannot be optimal since the cost can be made finite by choosing $q_R = 0$.

Assume that an optimal policy $\mathbf{q}$ does not have the structure stated in the Theorem. This would imply that q_R can be further decreased and q_G increased. In particular, one can perturb $\mathbf{q}$ in that way so that T_{acc} is unchanged. More precisely, note first that T_{acc} is monotone increasing in both q_R and in q_G. Hence

$$T_{acc}(1, q(R)) \geq T_{acc}(\mathbf{q}) \geq T_{acc}(q_G, 0),$$

Hence, if $T_{acc}(1, 0) < T_{acc}(\mathbf{q})$ then there is some $\mathbf{q_2}$ such that

$$T_{acc}(\mathbf{q_2}) = T_{acc}(\mathbf{q})$$

where

$$\mathbf{q_2} := (1, q_R^2), \quad \text{or} \quad \mathbf{q_2} := (q_G^2, 0) \quad \text{if } T_{acc}(1, 0) \geq T_{acc}(\mathbf{q})$$

We have

$$P_{\mathbf{q_2}}(I = 0) = 1 - T_{acc}(\mathbf{q_2}) = 1 - T_{acc}(\mathbf{q})$$

(e.g. from Little's Theorem). From rate balance arguments it follows that

$$P_{\mathbf{q_2}}(I = i) = (1 - T_{acc}(\mathbf{q_2}))\overline{\rho}_2^i \qquad \text{for } i \leq L. \tag{2.9}$$

Hence

$$P_{\mathbf{q_2}}(I \geq i) < P_{\mathbf{q}}(I \geq i) \tag{2.10}$$

for $i \geq L$. Thus

$$P_{\mathbf{q2}}(R) < P_{\mathbf{q}}(R).$$

By combining this with (2.2) it follows that

$$P_{\mathbf{q_2}}(I \geq i) = P_{\mathbf{q_2}}(R)\underline{\rho}(\mathbf{q_2})^{i-L} \leq P_{\mathbf{q_2}}(R)\underline{\rho}(\mathbf{q_2})^{i-L} \leq P_{\mathbf{q_2}}(I \geq i)$$

Hence (2.10) holds for all i. Taking the sum over i we thus obtain that

$$E_{\mathbf{q_2}}[I] < E_{\mathbf{q}}[I].$$

Since T_{acc} are the same under $\mathbf{q}$ and $\mathbf{q_2}$, it follows that $J_{\mathbf{q_2}} < J_{\mathbf{q}}$. Hence $\mathbf{q}$ is not optimal, which contradicts the assumption in the beginning of the proof. This establishes the structure of optimal policies. ■

2.4.2 Optimizing the Signal

Here we briefly discuss the case of choosing L so as to minimize $J_{\mathbf{q}}$ not only with respect to $\mathbf{q}$ but also with respect to the value L of the threshold.

To that end we first consider the problem of minimizing J over all stationary policies in case that full state information is available. This is a Markov decision process and an optimal policy is known to exists within the pure stationary policies. Moreover, a direct extension of the proof in Naor (1969) can be used to show that the structure of the optimal policy is of a threshold type: accept all arrivals as long as the state is below a threshold and reject all controlled arrivals otherwise. Note however that this policy makes use only of the information available also in our cases, i.e. of whether the state exceeds L or not.

We conclude that the problem of optimizing $J_{\mathbf{q}}$ over both L and $\mathbf{q}$ has an optimal pure threshold policy i.e. with $q_R = 0$ and $q_{G=1}$, or in other words $\mathbf{q} = (1, 0)$.

The optimal L for our problem can therefore be computed by minimizing $J_{\mathbf{q}}$ over pure threshold policies. Using Fig. 2.3, we compute this optimal L for $\mu = 1$, $\eta = 0.01$, $\lambda = 0.98$ and $\gamma = 1, 5, 10, 15, 20$ and obtain $L = 5$ for $\gamma = 20$.

2.5 The Game Problem

Note that the cost $J(q, P)$ corresponding to a strategy q of a player, when all others play P satisfies the following in order to be a best response to P: for each s, if $q(s)$ is not pure (is not 0 or 1) then at s, any other probability q' is also a best response.

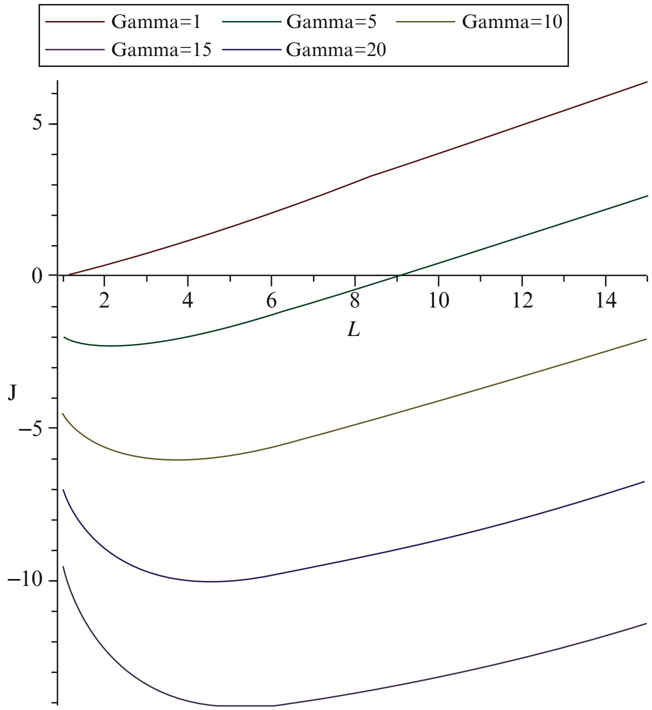

Fig. 2.3 The optimal performance for several values of L and γ

The cost for a user for entering when the signal is s given that the strategy of other users is $\mathbf{q} = (q_G, q_R)$ is given by

$$V_{\mathbf{q}}(s) = E_{\mathbf{q}}[W|s] - \gamma = E_{\mathbf{q}}[I|s]/\mu - \gamma \tag{2.11}$$

It is zero if it does not enter. Here, $E_{\mathbf{q}}[I|s]$ are given by

$$E_{\mathbf{q}}(I|R) = (L-1) + \frac{1}{(1-\underline{\rho})} \tag{2.12}$$

where $\mathbf{q} = (1, q_R)$ and where $\rho = (\nu + \zeta q_R)/\mu$, and

$$E_{\mathbf{q}}(I|G) = \frac{1}{1-\rho^L}\left(\frac{(\rho^L - \rho)}{\rho - 1} - (L-1)\rho^L\right) \tag{2.13}$$

where $\mathbf{q} = (q_G, 0)$ and where $\rho = (\nu + \zeta q_G)/\mu$. (The derivations of the above are as in (2.4) and (2.6), respectively.)

2.5.1 Structure of Equilibrium

The following gives the structure of equilibrium policies.

Theorem 2. *(1) The equilibrium policy is to enter for any signal if and only if* $V_{(1,1)}(R) \leq 0$.

(2) The equilibrium is of the form $\mathbf{q} = (1, q_R)$ *where* $q_R \in (0,1)$ *if and only if* $V_{(1,1)}(R) > 0 > V_{(1,0)}(R)$. *In this case, the equilibrium is given by the* $q = (1, q_R)$ *where* $\mathbf{q}_R$ *is the solution of* $V_{(1,q_R)}(R) = 0$ *where* $V_q(R)$ *is given in (2.11).*

(3) The equilibrium is of the form $\mathbf{q} = (q_G, 0)$ *where* $q_G \in (0,1)$ *if and only if* $V_{(1,0)}(G) < 0$. *In this case, the equilibrium is given by the* $\mathbf{q} = (q_G, 0)$ *where* q_G *is the solution of* $V_{(q_G,0)}(G) = 0$ *and where* $V_{\mathbf{q}}(G)$ *is given in (2.11).*

(4) The equilibrium policy is not to enter for any signal if and only if $V_{(0,0)}(G) \geq 0$.

(5) The equilibrium is (1,0) if and only if $V_{(1,0)}(R) \geq 0$ *and* $V_{(1,0)}(G) \leq 0$*).*

Note that if the condition $V_{(1,0)}(G) \geq 0$ in statement 3 holds then the condition of statement 2 does not hold since $V_{(1,0)}(G) \leq V_{(1,0)}(R)$.

Proof. We note that $E_{\mathbf{q}}(I|R) > E_{\mathbf{q}}(I|G)$ and hence

$$V_{\mathbf{q}}(R) > V_{\mathbf{q}}(G). \tag{2.14}$$

Assume $V_{(1,1)}(R) \leq 0$ then (1,1) is an equilibrium since no deviation of an arrival from (1,1) (i.e. always enter) can make its cost lower (since when a customer does not enter its cost is zero).

Let $\mathbf{q}$ be an equilibrium policy. Assume that $q_R > 0$. Then $V_{\mathbf{q}}(R) \leq 0$. Equation (2.14) implies that $V_{\mathbf{q}}(G) < 0$. Thus the following holds: (1) if (1,1) is an equilibrium then $V_{(\mathbf{1},\mathbf{1})}(s) \leq 0 \; s = R, G$. (2) At equilibrium, $q_G = 1$ and if $q_R < 1$ then by monotonicity in q_2 we have $V_{(\mathbf{1},\mathbf{1})}(R) > 0$ and $V_{(\mathbf{1},\mathbf{q})}(R) = V_{(\mathbf{1},\mathbf{q})}(G) = 0$. Conversely, if $V_{(1,1)}(R) > 0 > V_{(1,0)}(R)$ then by continuity we have some $q \in (0,1)$ for which $V_{(\mathbf{1},\mathbf{q})}(R) = 0$ and hence $(1, q)$ is an equilibrium: an arrival that receives a signal R is indifferent between joining or not the queue, whereas an arrival receiving a signal G has strict preference in joining the queue. This establishes the two first statements of the theorem.

Let again $\mathbf{q}$ be an equilibrium policy. Assume that $q_G < 1$. Then $V_{\mathbf{q}}(G) \geq 0$ so (2.14) implies that $V_{\mathbf{q}}(R) > 0$. Thus at equilibrium, $q_R = 0$. Thus the following holds: (1) if (0,0) is an equilibrium then $V_{(0,0)}(s) \geq 0 \; s = R, G$. (2) At equilibrium, $q_R = 0$ and if $q_G > 0$ then by monotonicity in q_1 we have $V_{(1,0)}(R) > 0$ and $V_{(q,0)}(R) = V_{(q,0)}(G) = 0$. Conversely, if $V_{(1,0)}(R) > 0 > V_{(0,0)}(R)$ then by continuity we have some $q \in (0,1)$ for which $V_{(q,0)}(R) = 0$ and hence $(q, 0)$ is an equilibrium: an arrival that receives a signal G is indifferent between joining or not the queue, where as an arrival receiving a signal R prefers strictly not to join the queue. This establishes statements 3–4 of the Theorem.

If the conditions in (5) hold then no user can benefit by entering the queue when the signal is red and no one can benefit by not entering when it is green. Hence (1,0) is an equilibrium. If the first condition of (5) is violated then deviating from (1,0) at a red signal is strictly better than not entering. If the second condition is violated then from (1,0) at a green signal and not entering the queue strictly improves the utility of a user. This proves statement 5. ■

2.5.2 Numerical Examples

We consider here as an example the parameters $\gamma = 20, \mu = 1, \lambda = 0.98 = \zeta$ and $\nu = 0.01$. For all L condition (1) of Theorem 2 does not hold, so (1,1) is not an equilibrium. Condition (2) of the Theorem holds for $L \leq 20$. In that case, the equilibrium is given by $(1, q_R)$ where q_R is given in Fig. 2.4. The value at equilibrium is given in Fig. 2.5 for the case of the signal G and is otherwise zero for all $L \leq 20$. For the case of $L > 20$ we have the opposite, i.e. $V_R = 0$. V_G is given by $E[I|G] - \gamma$ where $E[I|G]$ is expressed in (2.6).

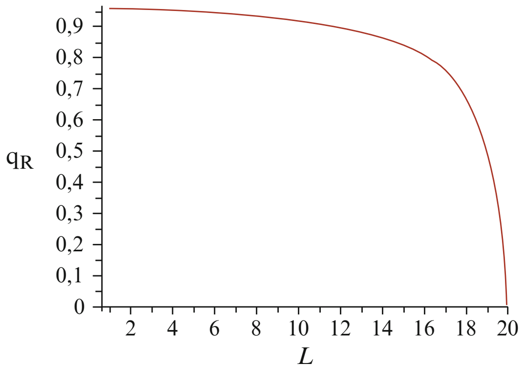

Fig. 2.4 Equilibrium action q_R as a function of L

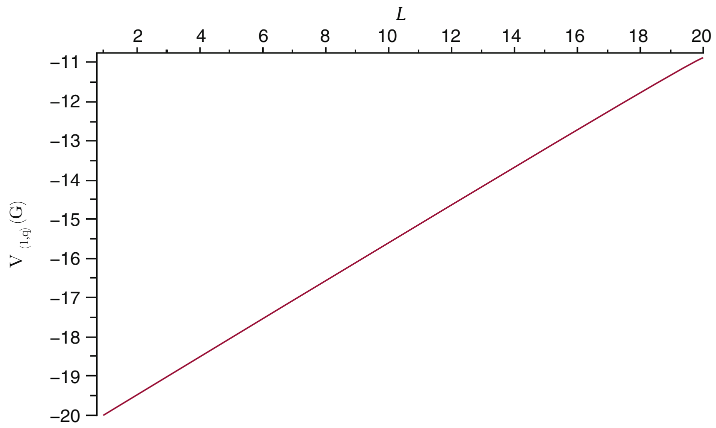

Fig. 2.5 Equilibrium value V_G for signal G as a function of L. We used case (2) in Theorem 2 and the results are therefore valid only for $L \leq 20$

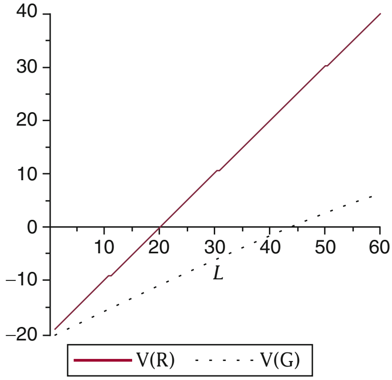

Fig. 2.6 $V_{\mathbf{q}}(R)$ and $V_{\mathbf{q}}(L)$ for $\mathbf{q} = (1, 0)$ as a function of L

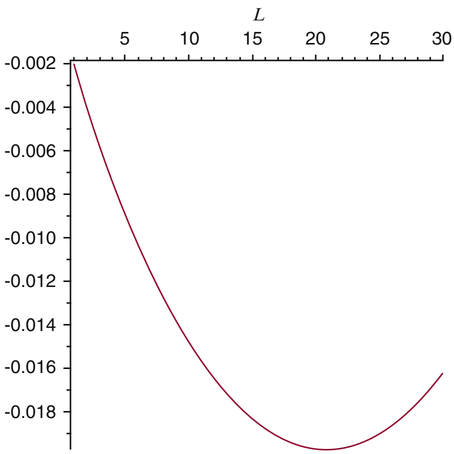

Fig. 2.7 The social value J at equilibrium as a function of L

Figure 2.6 depicts $V_{\mathbf{q}}(s)$ for $s = G, R$. We see that for L between 20 and 44, the conditions of statement 5 of Theorem 2 hold and thus the equilibrium is $\mathbf{q} = (1, 0)$ for this whole range of L.

Let L^* denote one plus the largest value L for which $V_{1,0} < 0$. L^* thus separates case (2) and (5) in Theorem 2. Then L^* equals the smallest integer greater than or equal to $\gamma\mu$. In our case it is given by 20 as is seen in Fig. 2.5. For every $L > L^*$ we know that, in fact, $q_R = 0$.

2.5.3 Optimizing the Signal

We are interested here in finding the L for which the induced equilibrium gives the best system performance. We plot the system performance J at equilibrium as a function of L in Fig. 2.7.

The optimal L is seen to equal 20 and the corresponding performance measures at equilibrium are $J = -0.02$ and $T_{acc} = 0.947$. We conclude that the policy for which the social cost is minimized has the same performance as the full state information equilibrium policy.

If we take the $L = 5$ which we had computed for optimizing the system performance, and use it in the game setting, we obtain $T_{acc} = 0.829$ and $J^* = -0.0088$ which indeed gives a much worse performance than the performance under the $L = 20$.

2.6 Concluding Remarks

We revisited in this paper the model introduced by P. Naor who studied the social and the individual optimal acceptance policy of arrivals at a queue. We studied a partially observable version in which arrivals are only informed on whether the queue length exceeds or not some given threshold value. We presented the structure of the optimal and the equilibrium policies and computed the value of L that leads to the best system performance at equilibrium.

Appendix: Uniform *f*-Geometric Ergodicity and the Continuity of the Markov Chain

The continuity of the steady state probabilities and thus of the expected queue length hold for the case of finite k since the chain is ergodic with finitely many states. We thus focus below on the case of infinite k. We show continuity of the expected queue length in $\mathbf{q}$ for q_R restricted to some closed interval for which the corresponding value of $\underline{\rho}_R$ is smaller than 1. (Due to Lemma 1 there exists indeed an interval such that any policy for which q_R is not in the interval cannot be optimal.)

We show that the Markov chain is f-Geometric Ergodic and then use Lemma 5.1 from Spieksma (1990).

Consider the Markov chain embedded at each transition in the queue size. Thus for $I \geq \max(L, 1)$, with probability β the event is a departure and otherwise it is an arrival, where

$$\beta := \frac{\mu}{\mu + \nu + q_R \zeta}.$$

Note $\alpha > 0$ implies that $\beta > 1/2$ (α is defined in the proof of Theorem 1).

Define $f(i) = \exp(\gamma i)$, for any $I \geq \max(L, 1)$,

$$E[f(I_{t+1}) - f(I_t)|I_t = i] = \beta \exp[\gamma(i-1)] + (1-\beta)\exp[\gamma(i+1)] - \exp(\gamma i)$$
$$= f(i)\Delta \quad \text{where} \quad \Delta = \beta z^{-1} + (1-\beta)z - 1$$

and where $z := \exp(-\gamma)$. Note that $\Delta = 0$ at

$$z_{1,2} = \frac{1 \pm \sqrt{1 - 4\beta(1-\beta)}}{2(1-\beta)} = \{1, \frac{\beta}{1-\beta}\}$$

Thus $\Delta < 0$ for all γ in the interval $\left(0, \log\left(\frac{\beta}{1-\beta}\right)\right)$ (which is nonempty since we showed that $1 > \beta > 1/2$). We conclude that for any γ in that interval, f is a Lyapunov function and the Markov chain is γ-geometrically ergodic uniformly in $\mathbf{q}$.

References

Altman E, Shimkin N (1998) Individual equilibrium and learning in processor sharing systems. Oper Res 46:776–784

Altman E, Gaujal B, Hordijk A (2000) Admission control in stochastic event graphs. IEEE Autom Control 45(5):854–867

Hordijk A, Spieksma F (1989) Constrained admission control to a queueing system. Adv Appl Probab 21:409–431

Hsiao MT, Lazar AA (1991) Optimal decentralized flow control of Markovian queueing networks with multiple controllers. Perform Eval 13:181–204

Korilis YA, Lazar A (1995) On the existence of equilibria in noncooperative optimal flow control. J ACM 42(3):584–613

Naor P (1969) On the regulation of queueing size by levying tolls. Econometrica 37:15–24

Spieksma F (1990) Geometrically ergodic Markov Chains and the optimal control of queues. PhD thesis, Leiden University

Stidham S (1985) Optimal control of admission to a queueing system. IEEE Trans Autom Control 30:705–713

Stidham S, Weber R (1989) Monotonic and insensitive optimal policies for control of queues with undiscounted costs. Oper Res 37:611–625

Stidham S, Rajagopal S, Kulkarni VG (1995) Optimal flow control of a stochastic fluid-flow system. IEEE J Sel Areas Commun 13:1219–1228

Yechiali U (1971) On optimal balking rules and toll charges in a $GI|M|1$ queueing process. Oper Res 19:349–370

Chapter 3
Methodological Issues in Analyzing Market Dynamics

Ariel Pakes

Abstract This paper investigates progress in the development of models capable of empirically analyzing the evolution of industries. It starts with a parallel between the development of empirical frameworks for static and dynamic analysis of industries: both adapted their frameworks from models taken from economic theory. The dynamic framework has had its successes: it led to developments that have enabled us to control for dynamic phenomena in static empirical models and to useful computational theory. However when important characteristics of industries were integrated into that framework it generated complexities which both hindered empirical work on dynamics per se, and made it unrealistic as a model of agent behavior. This paper suggests a simpler alternative paradigm, one which need not maintain all the traditional theoretical restrictions, but does maintain the core theoretical idea of optimizing subject to an information set. It then discusses estimation, computation, and an example within that paradigm.

Keywords Applied work on market dynamics • Markov perfect • Experience based equilibria

---

This paper is adapted from three lectures given in 2014: (1) The Cowles Lecture given at the North American Econometric Society, (in Minneapolis, June); (2) A Keynote Address at the International Society for Dynamic Games (in Amsterdam, July), and (3) A Keynote Address at the Latin American Econometric Society Meetings (in Sao Paulo, November). I have benefited from the comments of several participants in those conferences and from sets of detailed and very useful comments from Michael Dickstein, Liran Einav, Chaim Fershtman, Myrto Kalouptsidi and Florian Wagener. I wish to thank Oxford University Press for their kind permission to reproduce four tables that have first appeared in C. Fershtman and A. Pakes: "Dynamic games with asymmetric information: A framework for applied work", Quarterly Journal of Economics 127(4): 1611–1661, 2012.

A. Pakes (✉)
Thomas Professor of Economics at Harvard University Littauer 117, Cambridge, MA 02138, USA
e-mail: apakes@harvard.edu

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_3

MSC Codes: 62P20 (Game Theory), 35Q91 (Dynamical Systems), 58E17 (Statistics)

3.1 Introduction

It will be helpful if I start out with background on some recent methodological developments in empirical Industrial Organization, concentrating on those I have been more closely associated with. I start with an overview of what we have been trying to do and then move on to how far we have gotten. This will bring us naturally to the analysis of market dynamics, the main topic of the paper.

Broadly speaking the goal has been to develop and apply tools that enable us to better analyze market outcomes. The common thread in the recent developments is a focus on incorporating the institutional background into our empirical models that is needed to make sense of the data used in analyzing the issues of interest. These are typically the causes of historical events, or the likely responses to environmental and policy changes. In large part this was a response to prior developments in Industrial Organization theory which used simplified structures to illustrate how different phenomena could occur. The empirical literature was trying to use data and institutional knowledge to narrow the set of possible responses to environmental or policy changes (or the interpretations of past responses to such changes). The field was moving from a description of responses that could occur, to those that were "likely" to occur given what the data could tell us about appropriate functional forms, behavioral assumptions, and environmental conditions.

In pretty much every setting this required incorporating

- heterogeneity of various forms into our empirical models,

and, when analyzing market responses

- using equilibrium conditions to solve for variables that firms could change in response to the environmental change of interest.

The difficulties encountered in incorporating sufficient heterogeneity and/or using equilibrium conditions differed between what was generally labeled as "static" and "dynamic" models. For clarity I will use the textbook distinction between these two: (1) static models solve for profits conditional on state variables, and (2) dynamics analyzes the evolution of those state variables (and through that the evolution of market structure). By state variables here I mean: the characteristics of the products marketed, the determinants of costs, the distribution of consumer characteristics, the ownership structure, and any regulatory or other rules the agents must abide by. I begin with a brief review of the approach we have taken with static models, as that will make it easier to understand how the dynamic literature evolved.

Static Models

The empirical methodology for the static analysis typically relied on earlier work by our game theory colleagues for the analytic frameworks we used. The assumptions we took from our theory colleagues included the following:

- Each agent's actions affect all agents' payoffs, and
- At the "equilibrium" or "rest point"

  1. agents have "consistent" perceptions,[1] and
  2. each agent does the best they can conditional on their perceptions of nature's and their competitors' behavior.

Our contribution was the development of an ability to incorporate heterogeneity into the analysis and then adapt the framework to the richness of different real world institutions. This was greatly facilitated by progress in our computational abilities, and the related increased availability of data and econometric methodology. Of particular importance were econometric developments which enabled the use of semi-parametric (see Powell 1994) and simulation (see McFadden 1989; Pakes and Pollard 1989) techniques. The increased data, computing power and econometric techniques enabled the framework to be applied to a variety of industries using much weaker assumptions than had been used in the theory literature.

Indeed I would make the claim that the tools developed for the analysis of market allocations conditional on the state "variables" of the problem have passed a "market test" for success in an abundance of situations. I come to the conclusion for three reasons.

- First these tools have been incorporated into applied work in virtually all of economics that deals with market allocation when productivity and/or demand is part of the analysis.
- The tools are now used by public agencies, consultancies and to some extent by firms.
- The tools do surprisingly well, both in fit and in providing a deeper understanding of empirical phenomena.

For examples of the last point I note that empirical analysis of equilibrium pricing equations in retail markets that followed Berry et al. (1995) both (1) typically fit exceptionally well for a behavioral equation and (2) generated markups which were in accord with other sources of information on markups. Similarly the productivity analysis that followed Olley and Pakes (1996) was able to separate out and analyze changes in aggregate productivity attributable to (1) increases in productivity at the firm level and (2) increases resulting from reallocating output among differentially productive firms.

[1] Though the form in which the consistency condition was required to hold differed across applications.

I do not want to leave the impression that there is nothing left to be done in the analysis of equilibrium conditional on state variables. There have been several recent advances which have enhanced our ability to use static analysis to analyze important problems. This includes (1) the explicit incorporation of adverse selection and moral hazard into the analysis of insurance and capital markets (see e.g. Einav et al. 2012), (2) the analysis of upstream contracts in vertical markets characterized by bargaining (see Crawford and Yurukoglu 2013) and (3) the explicit incorporation of fixed (and other non-convex) costs into the analysis of when a good will be marketed (see Pakes et al. 2015).

Dynamic Models

Empirical work on dynamic models proceeded in a similar way to the way we proceeded in static analysis: we took the analytic framework from our theory colleagues and tried to incorporate the institutions that seemed necessary to analyze actual markets. The initial frameworks by our theory colleagues made assumptions which ensured that the

- state variables evolve as a Markov process,
- and the equilibrium was some form of Markov Perfection (no agent has an incentive to deviate at any value of the state variables).

In these models firms chose "dynamic controls": investments that determine the likely evolution of their state variables. Implicit in the second condition above is that players have perceptions of the controls' likely impact on the evolution of the state variables (their own and those of their competitors) and through that on their current and future profits, and that these perceptions are consistent with actual behavior (by nature, as well as by their competitors). The standard references here are Maskin and Tirole (1988a,b) for the equilibrium notion and Ericson and Pakes (1995) for the framework brought to applied work. Though, as we will see, there were a number of ways that this framework was successful, it has not had nearly the impact on empirical work that the static framework has, and I want to explore why.[2]

3.2 The Assumptions of the Dynamic Framework

We start by examining the two assumptions above in the context of symmetric information Markov Perfect models, the first dynamic models to be brought to data.

The Markov Assumption

Except in situations involving active experimentation and learning, where policies are transient, applied work is likely to stick with the assumption that states evolve as a (controlled) time homogeneous Markov process of finite order. There are a number

[2]There are, of course, some structural dynamic papers that are justifiably well known, see for example Benkard (2004), Collard-Wexler (2013), and Kalouptsidi (2014).

of reasons for this. First the Markov assumption is convenient and fits the data well in the sense that conditioning on a few past states (maybe more than one period in the past) is often all one needs to predict the controls. Second we can bound the gains from unilateral deviations from the Markov assumption (see Ifrach and Weintraub 2014), and have conditions which ensure those deviations can be made arbitrarily small by letting the length of the kept history grow (see White and Scherer 1994).

Finally, but perhaps most importantly, realism suggests information access and retention conditions as well as computational constraints limit the variables agents actually use in determining their strategies. I come back to this below, as precisely how we limit the memory has implications for the difference between the conditions empirical work can be expected to impose and those most theory models abide by.

Perfection

The type of rationality built into Markov Perfection is more questionable. It has clearly been put to good use by our theory colleagues, who have used it to explore possible dynamic outcomes in a structured way. It has also been put to good use as a guide to choosing covariates for empirical work which needed to condition on the impacts of dynamic phenomena [e.g. conditioning on the selection induced by exit in the analysis of productivity in Olley and Pakes (1996)]. However it has been less successful as an explicit empirical model of agents' choices that then combine to form the dynamic response of markets to changes in their environment. This because for many industries it became unwieldly when confronted with the task of incorporating the institutional background needed for an analysis of dynamic outcomes that many of us (including the relevant decision makers) would be willing to trust. The "unwieldliness" resulted from the dimension of the state space that seemed to be needed (this included at least Maskin and Tirole's, 1988, "payoff relevant" states, or the determinants of the demand and cost functions of each competitor), and the complexity of computing equilibrium policies. The difficulties with the Markov Perfect assumption became evident when we tried to use the Markov Perfect notions to structure

- the estimation of parameters, or to
- compute the fixed points that defined the equilibria or rest points of the system.

The initial computation of equilibrium policies in Pakes and McGuire (1994) discretized the state space and used a "synchronous" iterative procedure. The information in memory allowed the analyst to calculate policies and value functions for each possible state. An iteration circled through each state in turn and updated, first, their policies to maximize the expected discounted value of future net cash flow given the competitors' policies and the current firm's value function from the last iteration (i.e. it used "greedy" policies given the information in memory), and then the values the new policies implied. The test for equilibrium consisted of computing a metric in the difference between the values at successive iterations. If the metric was small enough we were close enough to a fixed point, and the fixed point satisfied all of the equilibrium conditions. The computational burden of this exercise varied directly with the cardinality of the discretized state space which grew (either exponentially or geometrically, depending on the problem) in the number of state variables.

At least if one were to use standard estimation techniques, estimation was even more computationally demanding, as it required a "nested fixed point" algorithm. For example a likelihood based estimation technique would require that the researcher compute equilibrium policies for each value of the parameter vector that the search algorithm tried in the process of finding the maximum to the likelihood. The actual number of equilibria that would need to be calculated before finding the maximum would depend on the problem but could be expected to be in the thousands.

The profession's initial response to the difficulties we encountered in using the Markov Perfect assumptions to structure empirical work was to keep the equilibrium notion and develop techniques to make it easier to circumvent the estimation and computational problems that the equilibrium notion generated. There were a number of useful contribution in this regard. Perhaps the most important of them were:

- The development of estimation techniques that circumvent the problem of repeatedly computing equilibria when estimating dynamic models (that do not require a nested fixed point algorithm). These used non-parametric estimates of the policy functions (Bajari et al. 2007), or the transition probabilities (Pakes et al. 2007), instead of the fixed point calculation, to obtain the continuation values generated by any particular value of the parameter vector.
- The use of approximations and functional forms for primitives which enabled us to compute equilibria quicker and/or with less memory requirements. There were a number of procedures used; Judd's (1998) book explained how to use deterministic approximation techniques, Pakes and McGuire (2001) showed how to use stochastic algorithms to alleviate the computational burden, and Doraszelski and Judd (2011) showed how the use of continuous time could simplify computation of continuation values.

As will be discussed in several places below many of the ideas underlying these developments are helpful in different contexts. Of particular interest to this paper, the new computational approaches led to an expansion of computational dynamic theory which illuminated several important applied problems. Examples include the relationship of collusion to consumer welfare (Fershtman and Pakes 2000), the multiplicity of possible equilibria in models with learning by doing (Besanko et al. 2010), and dynamic market responses to merger policy (Mermelstein et al. 2014). On the other hand these examples just sharpened the need for empirical work as the results they generated raised new, and potentially important, possible outcomes from the use of different policies, and we needed to determine when these outcomes were relevant. That empirical work remained hampered by the complexity of the analysis that seemed to be required were we to adequately approximate the institutional environment, at least if we continued to use the standard Markov Perfect notions.

3.2.1 The Behavioral Implications of Markov Perfection

I want to emphasize that the fact that the Markov Perfect framework becomes unwieldly when confronted by the complexity of real world institutions both limits our ability to do empirical analysis of market dynamics and raises the question of whether some other notion of equilibrium will better approximate agents' behavior. One relevant question then is: if we abandon Markov Perfection, can we both

- better approximate agents' behavior and,
- enlarge the set of dynamic questions we are able to analyze?

It is helpful to start by examining why the complexity issue arises. When we try to incorporate what seems to be the essential institutional background into our analysis we find that agents are required to both (1) access a large amount of information (all state variables), and (2) either compute or learn an unrealistic number of strategies (one for each information set). To see just how demanding this is consider markets where choices of consumers, as well as producers, have a dynamic component. This includes pretty much all markets for durable, experience and network goods—that is, it includes much of the economy.

In a symmetric information Markov Perfect equilibrium of, say, a durable good market, both consumers and producers would hold in memory at the very least (1) the Cartesian product of the current distribution of holdings of the good across households crossed with household characteristics, and (2) each firm's cost functions, both for the production of existing products and for the development of new products. Consumers would hold this information in memory, form a perception of likely product characteristics and prices of future offerings, and compute the solution to a sequential single agent dynamic programming problem to determine their choices. Firms would use the same state variables, take consumers decisions as given, and compute their equilibrium pricing and product development strategies. Since these strategies would not generally be consistent with the consumer's perceptions of those strategies that determined the consumers' decisions, the strategies would then have to be communicated back to consumers who would then have to recompute their value functions and choices based on the updated firm strategies. This process would need to be repeated until we found a "doubly nested" fixed point to the behavior of the agents; that is, until we found strategies where consumers do the best they can given correct perceptions of what producers would do and producers do the best they can given correct perceptions on what each consumer would do. It is hard to believe that this is as good an approximation to actual behavior as the social sciences can come up with.

A Theory "Fix"

One alternative to assuming that agents know all the information that would be required in a symmetric information Markov Perfect equilibrium is to assume agents only have access to a subset of the state variables. Since agents presumably know their own characteristics and these tend to be persistent, a realistic model would then need to allow for asymmetric information. In that case use of the "perfectness" notion would lead us to a "Bayesian" Markov Perfect solution. Though this will likely reduce information access and retention conditions, it causes a substantial increase in the burden of computing optimal strategies (by either the agents or the analyst). The additional burden results from the need to compute posteriors, as well as optimal policies, and the requirement that they be consistent with one another and hence with equilibrium strategies. The resulting computational burden would make it all but impossible to actually compute optimal policies (likely for many years to come). Of course there is the possibility that agents might learn these policies, or at least policies which maintain some of the logical features of Bayesian Perfect policies, from combining data on past behavior with market outcomes.

Learning Equilibrium Policies

Given its importance in justifying the use of equilibrium policies, there is surprisingly little empirical work on certain aspects of the learning process. There are at least three objects the firms need to accumulate information on: the primitives; the likely behavior of their competitors; and market outcomes given primitives, competitor behavior, and their own policies. There has been empirical work on learning about primitives,[3] but very little empirical (in contrast to lab experimental or theoretical) evidence on how firms formulate their perceptions about either their competitors' behavior, or about the impact of their own strategies given primitives and the actions of competitors.

An ongoing study by Doraszelski et al. (2014) delves into these questions. We study the British Electric Utility market for frequency response. Frequency response gives the Independent System Operator (a firm by the name of "National Grid") the ability to keep the frequency of the electricity network within regulated safety bounds. Until November 2005 frequency response was obtained by fiat through a regulation that required all units to allow National Grid to take control of a certain portion of their generating capacity. Starting in November 2005 a monthly auction market for frequency response replaced the regulatory requirement. We have data on bids, acceptances, and auxiliary information on this market from November 2005 until 2012. Note that when this market started the participants had no information available on either competitors' past bids, or about the response of price and quantities to the firms' own bids conditional on the competitors' bids. However they had dealt with the exogenous demand and supply characteristics of this market (monthly variation in demand, prices of fuel, etc.) for some time.

The results from that study which we are reasonably confident about and have relevance for this paper are that (1) the bids do eventually converge to what looks like an equilibrium, (2) after an initial stage where the learning process was too complex for our simple models to approximate adequately, bids for this good converge (and since the good is nearly homogeneous, there is a consequent dramatic fall in the inter-firm variance in bids), and (3) the many smaller changes in the environment thereafter do not seem to lead to further experimentation. Unfortunately I have

[3]See, for e.g. Crawford and Shum (2005), or for a recent contribution and a review of earlier work see Covert (2014).

little to say about modeling the periods of active experimentation that seem to have occurred in the period just after this market was formed. However I will come back below to the issue of learning models that do not involve experimentation.

I now turn to a notion of equilibrium that is less demanding than Markov Perfect for both the agents, and the analyst, to use. As we shall see many of the computational and estimation ideas that were developed for Markov Perfect models can be used with the new equilibrium notion, but new issues do arise. In particular, as is explained below, the notion of equilibrium that we propose admits a greater multiplicity than standard Markov Perfect models allow, so we will consider realistic ways of restricting the equilibria analyzed. The last section of the paper uses a computed example to explore associated computational issues.

3.3 Less Demanding Notions of Equilibria

I begin by considering conditions that would be natural candidates to characterize "rest points" of a dynamic system. I then consider a notion of equilibrium that satisfies those, and only those, conditions. The next subsection introduces an algorithm designed to compute policies that satisfy these equilibrium conditions. The algorithm can be interpreted as a learning process. So the computational algorithm could be used to model the response to a change in the industry's institutions, but only a change for which it is reasonable to model responses with a simple reinforcement learning process. In particular, I do not consider changes that lead to active experimentation.

Focusing on the equilibrium, or the rest point, makes the job of this subsection much easier, because strategies at a rest point likely satisfy a Nash condition of some sort; else someone has an incentive to deviate. However it still leaves open the question of the form and purview of the Nash condition. The conditions that I believe are natural and should be integrated into our modeling approach are that

- agents perceive that they are doing the best they can conditional on the information that they condition their actions on, and that
- if the information set that they condition on has been visited repeatedly, these perceptions are consistent with what they have observed in the past.

Notice that I am not assuming that agents form their perceptions in any "rational" (or other) way; just that they are consistent with what they have observed in the past, at least at conditioning sets that are observed repeatedly. Nor am I assuming that agents form perceptions of likely outcomes conditional on all information that they have access to. The caveat that the perceptions must only be consistent with past play at conditioning sets that are observed repeatedly allows firms to experiment when a new situation arises. It also implicitly assumes that at least some of the conditioning information sets are visited repeatedly; an assumption consistent with the finite state Markov assumption that was discussed above and which I will come back to below.

I view these as minimal conditions. It might be reasonable to assume more than this, for example that agents know and/or explore properties of outcomes of states not visited repeatedly. Alternatively it might the case that there is data on the industry of interest and the data indicate that behavior can be restricted further. I come back to both these possibilities after a more formal consideration of the implications of the two assumptions just listed.

Formalizing the Implications of Our Two Assumptions

Denote the information set of firm i in period t by $J_{i,t}$. The set $J_{i,t}$ will contain both public (ξ_t) and private ($\omega_{i,t}$) information, so $J_{i,t} = \{\xi_t, \omega_{i,t}\}$. The private information is often information on production costs or investment activity (and/or its outcomes). The public information varies considerably with the structure of the market. It can contain publicly observed exogenous processes (e.g. information on factor price and demand movements), past publicly observed choices made by participants (e.g. past prices), and whatever has been revealed over time on past values of $\omega_{i,t}$.

Firms chose their "controls" as a function of the information at their disposal, i.e. $J_{i,t}$. Typically potential entrants will decide whether or not to enter and incumbents will decide whether or not to remain active, and, if they remain active, how much to invest (in capital, R&D, advertising, etc.). Denote the policy chosen by firm i in period t by $m_{i,t} \in \mathcal{M}$, and for simplicity assume that the number of feasible actions, or $\#\mathcal{M}$, is finite [one can deal with continuous values of the control as do Ericson and Pakes (1995); see Fershtman and Pakes (2012)].

Also for simplicity assume we are investigating a game in which firms invest in their own privately observed state ($\omega_{i,t}$) and the outcomes depend only on the firms' own investment choices (not on the choices of their competitors).[4] In these games the evolution of the firm's own state is determined by a family of distributions which determine the likelihood of the firm's state in the next period conditional on its current state and the amount it invests, or

$$\mathscr{P}_\omega \equiv \{P(\cdot|\omega, m);\ \omega \in \Omega_\omega, m \in \mathcal{M}\}. \tag{3.1}$$

We assume the number of possible elements in Ω_ω, that is its cardinality (which will be denoted by $\#\Omega$), to be finite [though one can often derive this from primitives, see Ericson and Pakes (1995)].

[4]Though the outcomes could depend on the exogenous processes with just notational changes. This assumption, which generates games which are often referred to as capital accumulation games, is not necessary for either the definition of equilibrium, or the computational and estimation algorithms introduced below. Moreover, though it simplifies presentation considerably, there are many I.O. applications where it would be inappropriate. Consider, for example, a repeated procurement auction for, say timber, where the participants own lumber yards. Their state variable would include the fraction of their processing capacity that their current timber supply can satisfy. The control would be the bid, and the bids of others would then be a determinant of the evolution of their own state. For an analysis of these situations using the notion of equilibrium proposed here see Asker et al. (in process).

The firm's choice variables evolve as a function of $J_{i,t}$, and conditional on those choices, the private state evolves as a (controlled) Markov process. This implies that provided the public information evolves as a Markov process, the evolution of $J_{i,t}$ is Markov. In our computational example (Sect. 3.4), which is about maintenance decisions of electric utility generators, firms observe whether their competitors bid into the auction in each period (so the bids are public information), but the underlying cost "state" of the generator is private information and it evolves stochastically. Here I am simply going to assume that the public information, ξ_t, evolves as a Markov process on Ω_ξ and that $\#\Omega_\xi$ is finite.

In many cases (including our example) the finite state Markov assumption is not obvious. To derive it from primitives we would have to either put restrictions on the nature of the game [see the discussion in Fershtman and Pakes (2012)], or invoke "bounded rationality" type assumptions. I will come back to a more detailed discussion of this assumption below. This because the finite state Markov chain assumption is an assumption I need, and one that can be inconsistent with more demanding notions of equilibrium. For what is coming next one can either assume it was derived from a series of detailed assumptions, or just view it as an adequate approximation to the process generating the data.

Equation (3.1) and our assumption on the evolution of public information imply that $J_{i,t}$ evolves as a finite state Markov process on, say, $\mathscr{J}$, and that $\#\mathscr{J}$ is finite. Since agents choices and states are determined by their information sets, the "state" of the industry, which we label as s_t, is determined by the collection of information sets of the firms within it

$$s_t = \{J_{1,t}, \ldots, J_{n_t,t}\} \in \mathscr{S}.$$

If we assume that there are never more than a finite number of firms ever active [another assumption that can be derived from primitives, see Ericson and Pakes (1995)], the cardinality of $\mathscr{S}$, or $\#\mathscr{S}$, is also finite. This implies that any set of policies will ensure that s_t will wander into a recurrent subset of $\mathscr{S}$, say $\mathscr{R} \subset \mathscr{S}$, in finite time, and after that $s_{t+\tau} \in \mathscr{R}$ with probability one forever (Freedman 1971). The industry states that are in $\mathscr{R}$, and the transition probabilities among them, will be determined by the appropriate primitives and behavioral assumptions for the industry being studied.

For applied work, it is important to keep in mind that in this framework agents are not assumed to either know s_t or to be able to calculate policies for each of its possible realizations. Agent's policies (the exit and investment decisions of incumbents, and the entry and investment decisions of potentials entrants) are functions of their $J_{i,t} \in \mathscr{J}$ which is lower dimensional than $s_t \in \mathscr{S}$.

Back to Our Behavioral Assumptions

Our first assumption is that agents choose the policy (the $m \in \mathscr{M}$) that maximizes their own perception of the expected discounted value of future net cash flow. So we need notation for the agent's perceptions of the expected discounted value of future net cash flow that would result from the actions it could chose. The perception of the discounted value from the choice of policy m at state J_i will be denoted

$$W(m|J_i), \quad \forall m \in \mathscr{M}, \quad \forall J_i \in \mathscr{J}.$$

Our second assumption is that at least for a J_i that is visited repeatedly, that is for a J_i which is a component of an $s \in \mathscr{R}$, the agents' perceptions of these values are consistent with what they observe. So we have to consider what agents observe. When they are at J_i in period t they know the associated public information (our ξ_t) and observe the subsequent public information, or ξ_{t+1}. So provided they visit this state repeatedly they can compute the distribution of ξ_{t+1} given ξ_t. Assuming it is a capital accumulation game and that they know the actual physical relationship between investment and the probability distribution of outcomes (our $\mathscr{P}_\omega$), they can also construct the distribution of ω_{t+1} conditional on ω_t and m. Together this gives them the distribution of their next period's state, say J_i', conditional on J_i and m. Letting a superscript e denote an empirical distributions (adjusted for the impacts of different m), the conditional distributions are computed in the traditional way, that is by

$$\left\{ p^e(J_i'|J_i, m) \equiv \frac{p^e(J_i', J_i, m)}{p^e(J_i, m)} \right\}_{J_i', J_i}.$$

A firm at J_i which chooses policy m will also observe the profits it gets as a result of its choice. For simplicity we will assume that the profits are additively separable in m, as would occur for example if profits were a function of all observed firms' prices and m was an additive investment cost. Then once the firm observes the profits it obtains after choosing m it can calculate the profits it would have earned from choosing any $m \in \mathscr{M}$. The empirical distribution of the profits it earns from playing m then allows the firm to form an average profit from playing any m at J_i. We denote those average profits by

$$\pi^e(J_i|m), \quad \forall m \in \mathscr{M}, \quad \forall J_i \in \mathscr{J}.$$

Note that the profits that are realized at J_i when playing m depend on the policies of (in our example the prices chosen by) its competitors. These in turn depend on its competitors' states. In reality there will be a distribution of competitors' states, say J_{-i}, when the agent is at J_i, say

$$\left\{ p^e(J_{-i}|J_i) \equiv \frac{p^e(J_{-i}, J_i)}{p^e(J_i)} \right\}_{J_{-i}, J_i},$$

so in reality the actual expected profits of a firm who plays m at J_i is

$$\pi^e(J_i|m) = \sum_{J_{-i}} \pi(J_i, J_{-i}) p^e(J_{-i}|J_i) - m.$$

Given this notation, our two equilibrium conditions can be formalized as follows.

- If $m^*(J_i)$ is the policy chosen at J_i, our first equilibrium condition, i.e. that each agent chooses an action which maximizes its perception of its expected discounted value, is written as

$$W(m^*(J_i)|J_i) \geq W(m|J_i), \quad \forall m \in \mathscr{M}, \quad \forall J_i \in \mathscr{J}. \tag{3.2}$$

  Note that this is an equation which describes optimal choices of the agent. Note that provided the agent can learn the $\{W(\cdot|\cdot)\}$ (see below) the agent can make that choice without any information on the choices made by its competitors; the choice becomes analogous to that of an agent playing against nature.
- The second equilibrium condition is that for states that are visited repeatedly, i.e. that are in $\mathscr{R}$, these perceptions are consistent with observed outcomes. Since $W(m|J_i)$ is the perception of the expected discounted value of future net cash flows, for all m and for each J_i that is a component of an $s \in \mathscr{R}$ we require $W(m|J_i)$ to equal the average profit plus the discounted average continuation value where the distribution of future states needed for the continuation value is the empirical distribution of those states; that is, we require

$$W(m|J_i) = \pi^e(m|J_i) + \beta \sum_{J_i'} W(m^*(J_i')|J_i')p^e(J_i'|J_i). \tag{3.3}$$

Restricted Experience Based Equilibrium (or REBE)
The conditions in Eqs. (3.2) and (3.3) above are the conditions of a REBE as defined in Fershtman and Pakes (2012).[5] There also is related earlier work on "self-confirming" equilibria (see Fudenberg and Levine 1983) which is similar in spirit but differs in the conditions it imposes.

A Bayesian Perfect equilibrium satisfies the conditions of a REBE, but so do weaker notions of equilibrium. In particular the REBE does not restrict evaluations of states outside of the recurrent class to be consistent with the outcomes that play at those points would generate. As a result the REBE notion of equilibrium admits greater multiplicity than does Bayesian Perfect notions of equilibrium. We return to the multiplicity issue after explaining how to compute a REBE, as once one has the computational procedure clearly in mind, the multiplicity issue and ways of mitigating it can be explained in a transparent way.

The Equilibrium Conditions and Applied Work
We already noted that agents are not assumed to compute policies on, or even know, all of s_t; they only need policies conditional on J_i. Now note that there is nothing in our equilibrium conditions that forbids J_i from containing less variables than the decision maker has at its disposal. For example, if agents do not have the capacity to either store too much history or to form differing perceptions of expected

[5]In games where the agent can only use past data to calculate $\{\pi^e(J_i|m)\}$ for $m = m^*(J_i)$ and/or $p^e(J_i'|J_i, m)$ for $m = m^*(J)_i$, Fershtman and Pakes (2012) consider weakening the second condition to only require Eq. (3.3) to hold at $m = m^*(J_i)$. They call the equilibrium that results from the weaker notion an EBE (without the adjective 'restricted').

discounted values for information sets that detail too much history, one might think it is reasonable to restrict policies to be functions of a subset of the information available to the decision maker. This subset may be defined by a segment of the history, or a coarser partition of information from a given history. We come back to the question of how the analyst might determine the information sets that agents' policies condition on below.

The second point to note is related. There is nothing in these conditions that ensures that the policies we calculate on $\mathscr{R}$ satisfy all the equilibrium conditions typically assumed in the game theoretic literature. In particular it may well be the case that even if all its competitors formulated their policies as functions of a particular set of state variables, a particular firm could do better by formulating policies based on a larger set. For example in a model with asymmetric information it is often the case that, because all past history may be relevant for predicting the competitors' state, all past history will be helpful in determining current policies. In the absence of finite dimensional sufficient statistics, which for games are hard to find, this would violate the finite state Markov assumption on the evolution of public information. We could still, however, truncate the history and compute optimal policies for all agents conditional on the truncated history, and this would generate a Markov process with policies that satisfy our conditions (3.2) and (3.3).

Fershtman and Pakes (2012) discuss this in more detail; they consider alternative ways to ensure that REBE policies are the best an agent can do conditional on all agents forming policies as functions of the same underlying state space. Section 3.4 uses one of these for comparisons.[6] However I view the less restrictive nature of our conditions as an advantage of our "equilibrium" notion, as it allows agents to have limited memory and/or the ability to make computations, while it still imposes an appealing sense of rationality on the decision making process. Moreover in empirical work restrictions on the policy functions may be testable (see Sect. 3.3.2 below).

3.3.1 Computational Algorithm

The computational algorithm is a "reinforcement learning" algorithm,[7] similar to the algorithm introduced in Pakes and McGuire (2001). I begin by focusing on computational issues and consider the algorithm's behavioral interpretation thereafter.

---

[6]The example focuses on particular restrictions on the formation of policies, but there are many other ways of restricting policies which would generate Markov chains with similar properties. Indeed the papers I am aware of that compute "approximations" to Markov Perfect equilibria can be reinterpreted in this fashion; see for example Benkard et al. (2008) and Ifrach and Weintraub (2014), and the literature cited in those articles.

[7]For an introduction to reinforcement learning, see Sutton and Barto (1998).

From a computational point of view one of the algorithm's attractive features is that an increase in the number of variables that policies can be a function of—below, we will refer to these variables as "state" variables—does not (necessarily) increase the computational burden in an exponential, or even geometric, way.[8] Traditionally the burden of computing equilibria scales with both (1) the number of states at which policies must be computed, and with (2) the number of states we must integrate over in order to obtain continuation values. Depending on the details of the problem, both grow either geometrically or exponentially with the number of state variables, generating what is sometimes referred to as a "curse of dimensionality". The algorithm described below is designed to get around both these problems.

The algorithm is iterative and iterations will be indexed by a superscript k. It is also "asynchronous": each iteration only updates a single point in the state space. Thus an iteration has associated with it a location (a point in the state space), and certain objects in memory. The iterative procedure is defined by procedures for updating the location and the memory.

The location, say $L^k = (J_1^k, \dots J_{n(k)}^k) \in \mathscr{S}$, is defined as the collection of information sets of agents that are active at that iteration. The objects in memory, say M^k, include

1. a set of perceptions of the discounted value of taking action m at location J:

$$\mathscr{W}^k \equiv \{W^k(m|J_i), \forall m \in \mathscr{M} \text{ and } \forall J \in \mathscr{J}\},$$

2. a set consisting of the expected profits when taking action m at location J:

$$\Pi^k \equiv \{\pi^k(m|J_i), \forall m \in \mathscr{M} \text{ and } \forall J \in \mathscr{J}\},$$

3. the number of times each J has been visited prior to the current iteration, which we denote by h^k.

So the algorithm must update $(L^k, \mathscr{W}^k, \Pi^k, h^k)$.

Exactly how we structure and update the memory will determine the size of memory constraint and the compute time. Here I restrict myself to a structure that is easy to explain; the most efficient structure is likely to vary with the properties of the model and the computational facilities available. Also for clarity I work with a model with a specific specification for public and private information. I will assume that the private information, or ω, are payoff relevant states (e.g. costs of production), and the public information that is observed at any state is a function $b(m(J))$ of agents' controls. For instance, in the electric utility example computed below all agents see whether a generator is bid into the market, but only the owner of the generator sees whether maintenance is done on the generators not bid into the market. In addition the agent is assumed to know the primitive profit function

[8]The number of state variables in a problem is typically the number of firms that can be simultaneously active times the number of state variables of each firm.

$\pi(\cdot, b(m_{-i}))$, which can be used to compute counterfactual profits for any set $b(m_{-i})$ of competitors' controls; i.e. agents can compute $\pi(\cdot, m, b(m_{-i}))$ for $m \neq m_i^*$.

Updating the Location

The $\{W^k(m|J_{i,k})\}_m$ in memory represent the agent's perceptions of the expected discounted value of the future net cash flow that would result from choosing $m \in \mathcal{M}$. The agent chooses that value of m that maximizes these discounted values (it chooses the "greedy" policies). That is for each agent we choose

$$m_{i,k}^* = \arg\max_{m \in \mathcal{M}} W^k(m|J_{i,k}).$$

Next we take pseudo random draws on outcomes from the family of conditional probabilities in Eq. (3.1) conditional on $m_{i,k}^*$ and $\omega_{i,k} \in J_{i,k}$, that is from the family $P(\cdot|\omega_{i,k}, m_{i,k}^*)$. The outcomes from those draws determine $\omega_{i,k+1}$, which, together with the current bids of all agents (which is the additional public information), determine the $\{J_{i,k+1}\}$ and hence the new location L^{k+1}.

Updating the Memory

Updating the number of visits is done in the obvious way. I now describe the update of perceptions $(\Pi^k, \mathcal{W}^k)$. I do so in a way that accentuates the "learning" interpretation of the algorithm. Since we are using an asynchronous algorithm, each iteration only updates the memory associated with the initial location of that iteration.

We assume that the agent forms, for each hypothetical choice $m \in \mathcal{M}$, an ex post perception of what its profits and value would have been given the observed choices $b(m_{-i,k})$ made by other agents. These profits would have been found by evaluating the profit function at the alternative feasible policies conditional on its private state and its competitors' choice of policies, resulting in $\pi(\omega_{i,k}, m, b(m_{-i,k}))$. Similarly the value would have been those profits plus the continuation values that would have emanated from the alternative choices

$$V^{k+1}(J_{i,k}, m) = \pi(\omega_{i,k}, m, b(m_{-i,k})) + \max_{\tilde{m} \in M} \beta W^k(\tilde{m}|J_{i,k+1}(m)), \tag{3.4}$$

where $J_i^{k+1}(m)$ is what the time-$(k+1)$ information would have been, had the agent played m and had the competitors played their actual play. In the example this would require computing their returns from a counterfactual bid given the bids of the other agents.

The agent knows that $b(m_{-i,k})$ is only one of the possible actions its competitors might take when it is at $J_{i,k}$, as the actual action will depend on its competitors' private information, which the agent does not have access to. So it treats $V^{k+1}(J_{i,k}, m)$ as a random draw from the possible realizations of $W(m|J_{i,k})$, and updates $W^k(m|J_{i,k})$ by averaging this realization with those that have been generated from those prior iterations at which the agents' state was $J_{i,k}$. Formally

$$W^{k+1}(m|J_{i,k}) = \frac{1}{h^k(J_{i,k}) + 1} V^{k+1}(J_{i,k}, m) + \frac{h^k(J_{i,k})}{h^k(J_{i,k}) + 1} W^k(m|J_{i,k}),$$

or equivalently

$$W^{k+1}(m|J_{i,k}) - W^k(m|J_{i,k}) = \frac{1}{h^k(J_{i,k}) + 1}[V^{k+1}(J_{i,k}, m) - W^k(m|J_{i,k})].$$

An analogous formula is used to update expected profits $\{\pi^k(m|J_{i,k})\}$, i.e.

$$\pi^{k+1}(m|J_{i,k}) = \frac{1}{h^k(J_{i,k}) + 1}\pi(m|J_{i,k}, b(m(J_{-i,k}))) + \frac{h^k(J_{i,k})}{h^k(J_{i,k}) + 1}\pi^k(m|J_{i,k}).$$

This is a simple form of stochastic integration (see Robbins and Monro 1951). There are more efficient choices of weights for the averaging, as the earlier iterations contain less relevant information then the later iterations, but I do not pursue that further here.[9]

Properties of the Algorithm

Before moving to computational properties note that the algorithm has the interpretive advantage that it can be viewed as a learning process. That is agents (not only the analyst) could use the algorithm, or something very close to it, to learn equilibrium policies. This could be important for empirical work as it makes the algorithm a candidate tool for analyzing how agents might change their policies in reaction to a perturbation in their environment.[10]

We now consider computational properties of the algorithm. First note that if we had equilibrium valuations we would tend to stay there; i.e. if * designates equilibrium, then

$$E[V^*(J_i, m^*)|W^*] = W^*(m^*|J_i),$$

so there is a sense in which the equilibrium is a rest point to the system of stochastic difference equations. I do not know of a proof of convergence of reinforcement learning algorithms for (non zero-sum) games. However we provide a computationally convenient test for convergence below, and my experience is that the randomness in the outcomes of the algorithm together with the averaging over past outcomes that it uses typically is enough to overcome cycling problems that seem to be the most frequent (if not the only) manifestation of non-convergence.

[9]Except to note that the simple weights used here do satisfy Robbins and Monro's, 1951, conditions for convergence: the limit of the sum of the weights is infinite, while the limit of the sum of the squared weights is finite. Though those criteria do not ensure convergence in game theoretic situations, all applications I am aware of chose weights that satisfy them.

[10]Note, however, that were our algorithm to be used as tool for analyzing how agents react to a change in their environment one would have to clarify what information each agent has at its disposal when it updates its perceptions and modify the algorithm accordingly. That is in the algorithm described above we use all the information generated by the outcomes to all agents to update the perceptions of each agent, and this may not be possible in an actual application.

As noted algorithms for computing equilibria to dynamic games have two characteristics that generate computational burdens which increase rapidly as the number of state variables increases, and hence can generate a "curse of dimensionality". One is the increase of the number of points at which values and policies need to be calculated. In the algorithm just described the only states for which policies and values are updated repeatedly are the points in $\mathscr{R}$. The number of points in $\mathscr{R}$ need not increase in any particular way, indeed it need not increase at all, with the dimension of the state space. In the problems I have analyzed $\#\mathscr{R}$ does increase with the dimension of the state space, but at most linearly, rather than geometrically or exponentially (see the discussion in Sect. 3.4).

The second source of the "curse of dimensionality", as we increase the number of state variables, is the increase in the burden of computing the sum over possible future values needed to compute the continuation values at every point updated at each iteration. In this algorithm the update of continuation values is done as a sum of two numbers regardless of the number of state variables. Of course our estimate of continuation values involves simulation error while explicit integration does not. The simulation error is reduced by repeated visits to the point. The advantage of the simulation procedure is that the number of times a point must be visited to obtain a given level of precision in the continuation values does not depend on the dimension of the state space.

A computational burden of our algorithm that is not present in, say, the Pakes and McGuire (1994) algorithm, is that after finding a new location, the reinforcement learning algorithm has to search for the memory associated with that location. In traditional synchronous algorithms one simply cycles through the possible locations in a fixed order. The memory and search constraints typically only become problematic for problems in which the cardinality of $\mathscr{R}$ is quite large, and when they are problematic one can augment our algorithm to use functional form approximations such as those used in the "TD Learning" stochastic approximation literature (see Sutton and Barto 1998).

Convergence and Testing

Though the algorithm does not necessarily converge, Fershtman and Pakes (2012) provide a test for convergence whose computational burden is both small and independent of the dimension of the state space. To execute the test we first obtain a consistent estimate of $\mathscr{R}$. We then compute a weighted sum of squares of the percentage difference between (1) the actual expected discounted values from the alternative feasible policies and (2) our estimates of W at the points in $\mathscr{R}$. The weights are equal to the fraction of times the points in $\mathscr{R}$ would be visited, were those policies followed over a long period of time. This sum is an $L^2\big(P(\mathscr{R})\big)$ norm of the difference at the different points in $\mathscr{R}$, where $P(\mathscr{R})$ denotes the invariant measure on $\mathscr{R}$.

First note that any fixed estimate $\tilde{\mathscr{W}}$ of $\mathscr{W}$ generates policies which define a finite state Markov process for $\{s_t\}$[11]. To obtain a candidate for a recurrent class

[11] Formally we could gather the implied transition probabilities into a Markov matrix $Q(s', s|\tilde{\mathscr{W}})$ and describe our first step as finding a candidate for an $\mathscr{R}$ that is generated by $Q(s', s|\tilde{\mathscr{W}})$.

$\mathscr{R}(\tilde{\mathscr{W}})$ generated by those policies, start at any s, say s_0, and use the policies that $\tilde{\mathscr{W}}$ implies to simulate a sample path $\{s_t\}_{t=1}^{T_1+T_2}$. Let $\mathscr{R}(T_1, T_2, \cdot)$ be the set of states visited at least once between $t = T_1$ and $t = T_2$. This discards the points that are only visited during the first T_1 iterations of the algorithm, and keeps those that are visited between T_1 and T_2. Formally one can show that if $T_1, T_2 \to \infty$ in a way that ensures $T_2 - T_1 \to \infty$, the set $\mathscr{R}(T_1, T_2, \cdot)$ will converge to a recurrent class generated by the policies implied by $\tilde{\mathscr{W}}$. An operational way of checking whether a couple (T_1, T_2) of finite times is large enough is to continue simulating from T_2 to T_3, where, say, $T_3 - T_2 \approx T_2 - T_1$. Now check to see if the points visited between T_2 and T_3 are contained in $\mathscr{R}(T_1, T_2, \cdot)$.

Note that the policies we associate with $\tilde{\mathscr{W}}$ are optimal by construction; i.e. $m^*(J_i)$ is chosen to maximize $\{\tilde{W}(m|J_I)\}_{m\in\mathscr{M}}$. This brings us to our last equilibrium condition, the requirement that $\tilde{\mathscr{W}}$ is consistent with the actual outcomes from play for points in $\mathscr{R}$; i.e. we need to check whether

$$\tilde{W}(m|J_i) = \tilde{\pi}(m|J_i) \; + \beta \sum_{J_i'} \tilde{W}(m^*(J_i')|J_i')p^e(J_i'|J_i), \quad \forall m \in \mathscr{M}, \quad J_i \subset s \in \mathscr{R},$$

where $\tilde{\pi}(m|J_i)$ is the algorithm's estimate of expected profits.

In principle we could check this condition by direct summation, but that would be computationally burdensome (indeed it would bring the curse of dimensionality back into play). So we now show how to use simulated sample paths to check it. Start at an $s_0 \in \mathscr{R}$ and use the policies generated by $\tilde{\mathscr{W}}$ to forward simulate. At each J_i visited compute perceived values; that is, compute $V^{k+1}(\cdot)$ as in Eq. (3.4). Since we are simulating a recurrent process on its recurrent class the simulation run will visit each J_i in $\mathscr{R}$ repeatedly. Keep track of the average and the sample variance of the simulated perceived values at each point, say

$$\left(\hat{\mu}\left(\tilde{W}(m(J_i)|J_i)\right), \hat{\sigma}^2\left(\tilde{W}(m(J_i)|J_i)\right)\right).$$

Let $E(\cdot)$ take expectations over the simulated random draws and, for expositional simplicity, omit the index i. Then note that we can compute

$$\begin{aligned}
\mathscr{T}_{m,J} &\equiv E\left(\frac{\hat{\mu}(\tilde{W}_{m,J}) - \tilde{W}_{m,J}}{\tilde{W}_{m,J}}\right)^2 \\
&= E\left(\frac{\hat{\mu}(\tilde{W}_{m,J}) - E[\hat{\mu}(\tilde{W}_{m,J})]}{\tilde{W}_{m,J}}\right)^2 + \left(\frac{E\left[\hat{\mu}(\tilde{W}_{m,J})\right] - \tilde{W}_{m,J}}{\tilde{W}_{m,J}}\right)^2 \\
&= \%\,\mathrm{Var}(\hat{\mu}(\tilde{W}_{m,J})) \quad + \quad \%\,\mathrm{Bias}^2(\hat{\mu}(\tilde{W}_{m,J})).
\end{aligned}$$

The quantity $\mathscr{T}_{m,J}$ is the percentage mean square error in our estimate of the expected discounted value of taking action m when at state J; i.e. it is the sum of the percentage bias and the percentage variance of the estimate.

Let

$$\mathscr{T}_J \equiv M^{-1} \sum_{m \in \mathscr{M}} \mathscr{T}_{m,J},$$

where $M = \#\mathscr{M}$. Then $\mathscr{T}_J$ is the average percentage mean square error in the evaluation of the actions that can be taken when at J. As the number of simulation draws grows, the law of large numbers implies that we can obtain a consistent estimate of the contribution of the variance in the sample paths to $\mathscr{T}_J$. That is

$$\sum_J f_J \frac{1}{M} \sum_{m \in \mathscr{M}} \left(\frac{\hat{\sigma}^2(\tilde{W}_{m,J})}{\tilde{W}^2_{m,J}} \right) - \sum_J f_J \frac{1}{M} \sum_{m \in \mathscr{M}} \left(\frac{\hat{\mu}(\tilde{W}_{m,J}) - E[\hat{\mu}(\tilde{W}_{m,J})]}{\tilde{W}_{m,J}} \right)^2 \stackrel{a.s.}{\rightarrow} 0.$$

Consequently if

$$\text{Bias}(\mathscr{W}_{\mathscr{R}}) \equiv \sum_J f_J \mathscr{T}_J - \sum_l f_J \frac{1}{M} \sum_{m \in \mathscr{M}} \left(\frac{\hat{\sigma}^2(\tilde{W}_{m,J})}{\tilde{W}^2_{m,J}} \right),$$

then

$$\text{Bias}(\mathscr{W}_{\mathscr{R}}) \stackrel{a.s.}{\rightarrow} \sum_J f_J \frac{1}{M} \sum_{m \in \mathscr{M}} \left(\frac{E\Big[\hat{\mu}(\tilde{W}_{m,J})\Big] - \tilde{W}_{m,J}}{\tilde{W}_{m,J}} \right)^2,$$

which is an $L^2(\mathscr{P}_{\mathscr{R}})$ norm in the percentage bias, where $\mathscr{P}_{\mathscr{R}}$ is the invariant measure associated with $(\mathscr{R}, \tilde{W})$.

If $\text{Bias}(\mathscr{W}_{\mathscr{R}})$ is zero and $\mathscr{R}$ is a recurrent class then all of our equilibrium conditions are satisfied. Notice that this test statistic has an easy interpretation: it is the percentage difference between our estimate of and the actual expected discounted value of the net cash flow from the policies that can be undertaken from points in the recurrent class. The test is integrated into the computational algorithm by calling it after every fixed number of iterations, and stopping the algorithm when the estimate of $\text{Bias}(\mathscr{W}_{\mathscr{R}})$ is sufficiently small.

3.3.2 Empirical Challenges and Estimation

I am going to assume that the static profit function is known, as there has been a large literature devoted to empirically analyzing its components [see for instance the first two sections of Ackerberg et al. (2007), and the literature cited there].

The empirical researcher will still need to determine J_i and possibly estimate "dynamic" parameters, that is, parameters that are not determinants of the static profit function.

Determining J_i

In most empirical work the authors simply assume knowledge of J_i, or of the arguments of the policy functions. However given that part of our motivation is to reduce the complexity of the problem by limiting the content of J_i, some discussion of how to determine J_i is in order.

The first thing to note is that what we need to find out is the determinants of the dynamic controls (investment, entry, and exit in our example). In particular it may well be the case that decision makers do not condition on all the information available to them in making these decisions, possibly because making predictions for too fine a partition of the state space is too complicated. As a result specifying a J_i that includes all the information we know the decision maker has access to may not be necessary or even appropriate.

This suggests two, hopefully reinforcing, methods of determining J_i. The first is an empirical analysis of the determinants of the dynamic controls. The second, which may not always be possible, is to ask decision makers from the industry about which information their decisions on the dynamic controls depends (see for e.g. Wollmann 2015). There are likely to be two sources of error or disturbances in our predictions for the dynamic controls (1) a "structural" disturbance which results from a determinant of the agent's choice that we do not observe and (2) a disturbance due to measurement error. Ideally the structural error would be independently distributed over time. and the measurement error component should not be correlated with variables which are thought to be correctly measured. As a result a test of whether the disturbance we obtain from our predictions for the controls satisfies these ideal conditions is that they be uncorrelated with—actually independent of—past values of correctly measured variables. If there is an indication that the disturbances have a noticeable correlation with past values of correctly measured variables, one should allow for a serially correlated unobserved state variable (see below for further discussion).

Estimating Dynamic Parameters

The estimates needed will be obtained from firm or establishment level data (depending on the parameters being estimated). As a result they will often be based on data sets of similar size as the data sets used in estimating "static" models. These are frequently large enough to obtain reasonably precise parameter estimates (see Sect. 3.1).

Typically many of the dynamic parameters can be estimated by careful analysis of the relationship between observables without using any of the constructs that need to be computed from the equilibrium to the dynamic model (such as expected discounted values). For example if investment (our control, m) is observed and directed at improving a measure of a stock of some form (our ω), and the stock is either observable or can be backed out of the profit function analysis, the parameters of $P(\cdot|m, \omega)$ can be estimated directly from the relationship between m and ω's in adjacent periods. However there often are some parameters that can only be estimated through their relationship to perceived discounted values; sunk and fixed costs often have this feature. Also, where possible, more efficient estimators of dynamic parameters that can be estimated without using discounted values can be obtained by using these values.

There is a review of the literature on estimating parameters using the implications of the dynamic model in the third section of Ackerberg et al. (2007). That review focuses on symmetric information Markov Perfect models and emphasizes the tradeoff between statistical efficiency (in the sense of lower asymptotic variance of an estimated parameter), and computational efficiency (or the computational burden of the estimator), in the choice of estimators. It assumes that the state variables of the problem are known to the analyst and provides details on estimators which use them but avoid nested fixed point algorithms.[12] These estimators are all two-step estimators. The first step obtains non-parametric estimates of either (1) the probabilities of various actions, (i.e. of the "dynamic" controls), as functions of the state variables of the model—typically this includes the probability of entry and exit and a distribution for investment policies; see Bajari et al. (2007)—or (2) direct estimates of the Markov transition matrix for the state variables derived from those policies (Pakes et al. 2007). The second step then uses the transitions implied by the non-parametric estimates and the profit function to compute the discounted value of alternative actions conditional on the parameter of interest. It then finds that value for the parameter vector which makes the prediction for the optimal value of the control as close as possible to the choices actually made for that control.

For example given the profit function, the evolution of the state variables, and the probabilities of exit at each state, we can compute the expected discounted value of an entrant in any period. If the model is correct and we observe entry, the expected discounted value generated by entering should be higher than the sunk cost of entry, whereas if we do not observe entry, this expected discounted value should be lower than those costs. Since the average of the realized discounted values should approximate the average of the expected discounted values, the average of the discounted values in the periods when we do, and when we do not, observe entry can be used to estimate bounds on the sunk cost of entry. At the cost of a slight increase in computational burden, one can incorporate heterogeneity in sunk costs and use point (instead of set) estimators in these models (see Ackerberg et al. 2007). Given J_i, estimators for Markov Perfect models with asymmetric information can be computed analogously.

In addition I now describe a "perturbation" estimator, similar to the Euler equation estimator for single agent dynamic problems proposed by Hansen and Singleton (1982). This estimator does not require the first step non-parametric estimator, and can be used for estimation in models with asymmetric information; these estimators are not available for symmetric information Markov Perfect models, see below.

[12]Nested fixed point estimators are estimators that require the analyst to compute a new equilibrium every time one evaluates a different parameter vector in the estimation algorithm; see Rust (1994).

The perturbation estimator uses the inequality condition $W(m^*|J_i) \geq W(m|J_i)$ for equilibrium policies to generate (set) estimators of parameters. As in the literature on estimating parameters from symmetric information dynamic models, we assume that information on the equilibrium values of controls chosen on the recurrent states are available from past play.

Recall that J_i contains both public and private information. Let J^1 have the same public information as J^2, but different private information. If a firm is at J^1 it knows it could have played $m^*(J^2)$ and its competitors would respond by playing *on the equilibrium path* from J^2.[13] If J^2 is in the recurrent class we will have data on what competitors would have done were the agent to have chosen $m^*(J^2)$. Provided that this choice results in outcomes in $\mathscr{R}$, we can simulate a sample path from J^2 using only observed data on equilibrium play in $\mathscr{R}$. The Markov property ensures that the simulated path starting from the deviation to $m^*(J^2)$ will intersect the actual observed sample path at a random stopping time with probability one. From that time forward the two paths would generate the same profits. So the difference in discounted net cash flow from the sequence starting at the actual $m^*(J^1)$ and the sequence starting from the deviation, i.e. from $m^*(J^2)$, is just the difference in discounted returns from the period of the deviation to the time when the paths meet; this difference we can calculate. Since the initial choice of $m^*(J^1)$ was optimal, the conditional expectation of the difference in discounted profits between the simulated and actual path from the period of the deviation to the random stopping time, should, when evaluated at the true parameter vector, be positive. This yields moment inequalities for estimation as in Pakes et al. (2015), or the alternatives noted in Pakes (2010).

As noted it may well be important to integrate serially correlated unobservables into these estimation routines. Integrating serially correlated unobservables into these procedures can raise additional issues, particularly if the choice set is discrete. There has been recent work on discrete choice models that allow for serially correlated unobservables (see Arcidiano and Miller 2011; Pakes and Porter 2014), but it has yet to be used in problems that involve estimating parameters that determine market dynamics.

3.3.3 Multiplicity of Equilibrium Policies

We noted in Sect. 3.3 that REBE conditions admit more equilibria than Bayesian Perfect conditions. To see why, partition the points in $\mathscr{R}$ into "interior" and "boundary" points.[14] Points in $\mathscr{R}$ at which there are feasible (but inoptimal)

---

[13]It is the fact that data would not tell us the response to off the equilibrium path behavior for symmetric information Markov Perfect models that makes the perturbation technique inappropriate for estimating parameters based on those models.

[14]This partitioning is introduced in Pakes and McGuire (2001).

strategies which can lead outside of $\mathscr{R}$ are boundary points. Interior points are points that can only transit to other points in $\mathscr{R}$ no matter which of the feasible policies are chosen.

Our conditions only ensure that perceptions of outcomes are consistent with the results from actual play at interior points. Perceptions of outcomes for feasible (but non-optimal) policies at boundary points need not be tied down by actual outcomes. As a result differing perceptions of discounted values at points outside of the recurrent class can support different equilibria. This is a major reason for the existence of REBE which are not Bayesian Perfect.[15]

One can mitigate the multiplicity problem by adding either empirical information or by strengthening the behavioral assumptions. Without going into details we note that in an empirical application the data will contain information on which equilibrium has been played. For example, if J_i and m^* are observable we will know policies for states in $\mathscr{R}$. This in turn implies inequalities on the equilibrium $\{W(m|\cdot)\}$ which should rule out some equilibria. Moreover if profits are observed or estimated they can be used, together the transition probabilities, to directly compute estimators of $\{W(m|\cdot)\}$ by simulating sample paths. This may well eliminate other equilibria.[16]

These sources of information will be less helpful in two important cases: (1) when attempting to analyze counterfactuals, as we often want to do when examining the impacts of policy or environmental changes,[17] or when (2) computing equilibria in cases where we are willing to specify primitives but do not have historical data. In these and other cases where we need to augment whatever empirical information is available on the choice of equilibrium, it may be reasonable to invoke stronger behavioral assumptions. One possibility is to invoke learning rules, like the one in our algorithm, and simulate equilibria using one or more such rules. This is likely to be more helpful in analyzing counterfactual perturbations to a known environment, as then there is a natural initial condition, the current state of the industry, to start the learning process from. A second possibility is to impose additional restrictions on the equilibrium concept per se. I turn to this possibility now.

[15]There are other reasons for differences between REBE and Bayes Nash equilibria. For example, as noted above we do not assume that agents necessarily base their decisions on all the information they either have, or could have, access to. Also typically Bayes Nash Equilibria are defined in terms of consistency of perceived probability distributions with actual actions, whereas we are defining the equilibria in terms of consistency of perceived expected values with actual realized values. There can be different probability distributions that lead to the same expectation. Assuming agents wish to maximize expected discounted value they should be indifferent between two distributions with the same expectations, so I do not see this difference as substantive.

[16]However I know of no formal work which provides details on the extent to which the information in a particular data set limits the set of equilibria that could have generated it.

[17]Assuming historical data is available, there are two different cases here: one is a counterfactual which changes the underlying state space, and one that does not. If the state space is unchanged *and* one assumes that the counterfactual does not change the equilibrium selection mechanism, it would be possible to use historical data to guide the choice of the counterfactual equilibrium.

In many cases prior knowledge or past experimentation will endow agents with realistic perceptions of the value of states outside, but close to, the recurrent class. In these cases we will want to impose conditions that ensure that the equilibria we compute are consistent with this knowledge. To accommodate this possibility, Asker et al. (2014) propose an additional condition on equilibrium play that ensures that agents' perceptions of the outcomes from all feasible actions at points in the recurrent class are consistent with the outcomes that those actions would generate. They label the new condition "boundary consistency" and provide a computational simple test to determine whether the boundary consistency condition is satisfied for a given set of policies. We now formalize that condition.

Let $B(J_i|\mathscr{W})$ be a set of actions at J_i, which is a component of $s \in \mathscr{R}$, that could generate outcomes which are not in the recurrent class (so J_i is a boundary point), and let $B(\mathscr{W}) = \cup_{J_i \in \mathscr{R}} B(J_i|\mathscr{W})$ be the set of all possible actions of this kind. Then the notion of "boundary consistency" is formulated as follows.

Boundary Consistency Let τ index future periods, and consider a fixed estimate $\tilde{\mathscr{W}}$ of $\mathscr{W}$. Then $\tilde{\mathscr{W}}$ generates boundary consistent policies if for all $(m, J) \in B(\tilde{\mathscr{W}})$

$$E\left[\pi(m_i, J_i, J_{-i,0}) + \sum_{\tau=1}^{\infty} \delta^{\tau} \pi\Big(m(J_{i,\tau}), m(J_{-i,\tau})\Big)\Big| m_i = m, J_i = J, \tilde{\mathscr{W}}\right] \leq \tilde{W}(m^*|J),$$

where $E[\cdot|m_i = m, J_i = J, \tilde{\mathscr{W}}]$ takes expectations over the current states of the competitors, and the future states of all firms, conditional on $J_i = J$ and $m_i = m$, using the policies generated by $\tilde{\mathscr{W}}$. ♠

The boundary consistency condition ensures that the policies chosen at the boundary points yield higher discounted values than those of other feasible actions if competitors follow the policies in memory at all states, including the states not in the recurrent class but that communicate with a boundary point if some feasible policy is taken.

The test for boundary consistency uses the fact that we have $\mathscr{W}$ estimates in memory for points outside of $\mathscr{R}$. It uses these $\mathscr{W}$ to determine policies at those points and then to simulate sample paths from each (m, J_i) in $B(\tilde{\mathscr{W}})$. The null hypothesis states that the values of all sample paths from feasible but non-optimal policies are less than the values of sample paths from optimal play from the same state. Asker et al. (2014), show how to formulate a test of this null.[18] The test should rule out equilibria that are supported by perceptions of play outside of $\mathscr{R}$ that are unrealistic in the sense that they do not accord with the profits to be earned

[18]The test statistic is formed by taking a weighted average of the positive parts of the difference between the estimated value of feasible play and of optimal play of the states in $B(\mathscr{W})$, normalized by the variance of that difference. Since this is a statistic formed from moment inequalities (in contrast to moment equalities), the distribution of this statistic does not have a pivotal form and so needs to be simulated. However the critical values for it are relatively easy to simulate and are compared to the actual value of the test statistic to determine whether to accept the null (for details see Asker et al. 2014).

in those locations. The accuracy of the estimates of the sample paths from boundary points will depend on the components of $\mathscr{W}$ that are not associated with points in $\mathscr{R}$. However they are connected to $\mathscr{R}$ through feasible play, and hence may well have been explored both by agents and by the computational algorithm we use to compute equilibria.

Ergodicity
There is another type of multiplicity that may be encountered as there may be multiple recurrent classes for a given equilibrium policy vector. A sufficient condition for the policies to generate an ergodic process—a process with a unique recurrent class—is that there is a single state which can be reached from all states (Freedman 1971). Ericson and Pakes (1995) use this condition together with assumptions on primitives to prove ergodicity for a certain class of Markov Perfect models. However, in our notation those conditions would be a function of $\mathscr{W}$ on all of $\mathscr{S}$, and our estimates of the W at points not in $\mathscr{R}$ are imprecise (which would make it difficult to determine if those conditions are satisfied). Moreover there are cases of interest where multiple separate recurrent classes are likely (see Besanko et al. 2014). Of course data on, or even qualitative knowledge of, industry structures should help us pick out which (if there are many) recurrent class is appropriate for the problem at hand.

3.4 Computational Results from an Example

It is easiest to explain the computational issues in the context of an example, so most of my focus will be on the example in Fershtman and Pakes (2012) which is concerned with the maintenance decisions of electric utility generators.

The restructuring of electricity markets has focused attention on the design of markets for electricity generation. One issue in this literature is whether the market design would allow generators to make super-normal profits during periods of high demand. In particular the worry is that the twin facts that currently electricity is not storable and has extremely inelastic demand might lead to sharp price increases in periods of high demand.

The analysis of the sources of price increases during periods of high demand typically conditions on whether or not generators are bid into or withheld from the market. Generators have to go down for maintenance periodically. Since the benefits from incurring maintenance costs today depend on the returns from bidding the generator in the future, and the latter depend on what the firms' competitors bid at future dates, an equilibrium framework for analyzing maintenance decisions requires a dynamic game with strategic interaction. The Fershtman-Pakes paper provides a simple example of a REBE to a game that endogenizes maintenance decisions.

Table 3.1 Model details

Parameter	Firm L	Firm S
Number of generators	2	3
Range of ω	0–4	0–4
MC @ $\omega = (0, 1, 2, 3)$[a]	(20,60,80,100)	(50,100,130,170)
Capacity at constant MC	25	15
Costs of maintenance	5000	2000

[a]MC is constant at this cost until capacity and then goes up linearly. At $\omega = 4$ the generator shuts down

Details of the Model

The model has two firms. Firm "S" has three small generators with low start up costs and high marginal costs (they represent "gas fired" generators), and firm "L" has two large generators with high start up costs and low marginal costs (they represent "coal fired" generators); see Table 3.1. Each generator can bid supply functions into an independent system operator (the ISO). The ISO sums the bid functions, and then intersects the resultant supply curve with demand (which varies by day of the week) to determine a price. That price is paid to all electricity bid in at any price below it (this is a uniform price auction). The generators have constant marginal cost until the capacity listed in Table 3.1 after which marginal costs are increasing. Each generator also has a productivity variable, our ω, which is a private information state variable; productivity decays stochastically with use. The demand curve is inelastic.

Each period the firm choses among three actions for each of its generators. It can

- bid the generator into the market, which we denote by $m = 2$,
- withhold the generator from the market and use the period to do maintenance on the generator, our $m = 1$, or
- withhold the generator from the market and do not do any maintenance, which we denote by $m = 0$.

If the generator is bid in we assume, for simplicity, that it always bids in the same supply curve: so the firm's bid function is $b(m_i) : m_i \rightarrow \{0, b_i\}^{n_i}$, where b_i is the fixed bid schedule for the generators of firm i, and firm i has n_i generators. Firms do not see whether their competitors do maintenance but they do see their competitors' bids. So m_i is not in the public information, but $b(m_i)$ is.

The cost of producing electricity on each firm's generators is private information; it is a function of the productivity of the generator (our $\omega \in \Omega$) and the quantity of electricity the generator produces ($q_{i,t}$). So the cost function, our $c_i(\omega_{i,t}, q_{i,t})$, is increasing in both its arguments. The state ω increases stochastically with use, but reverts to a starting value if the firm does maintenance. If generators are indexed by r, formally

- $m_{i,r,t} = 0$ implies $\omega_{i,r,t+1} = \omega_{i,r,t}$,
- $m_{i,r,t} = 1$ implies $\omega_{i,r,t+1} = \overline{\omega}_{i,r}$ where $\overline{\omega}$ is the starting value, and
- $m_{i,r,t} = 2$ implies $\omega_{i,r,t+1} = \omega_{i,r,t} + \eta_{i,r,t}$ where η is a random "productivity shock" with $P(\eta) > 0 \; for \;\; \eta \in \{0, 1\}$.

If d is demand on that day and f is maintenance cost (our "investment"), the price p is determined by a function $p = p(b(m_i), b(m_{-i}), d)$ and $q = q(b(m_i), b(m_{-i}), d)$ is the allocated quantity vector. Realized profits for firm i are the sum of its profits from its generators minus the cost of maintenance, or

$$\pi_i(\omega_i, m_i, b(m_{-i}), d) \equiv \sum_r p_t q_{i,r,t} - \sum_r c_i(\omega_{i,r,t}, q_{i,r,t}) - f_i \sum_r \{m_{i,r,t} = 1\}.$$

In this game the bid function $b(m)$ is the only signal sent in each period. In particular $b(m_{-i,t-1})$ is a signal on $\omega_{-i,t-1}$, which cannot be observed by i, but which is a determinant of $b(m_{-i,t})$, and consequently of $\pi_{i,t}$.

3.4.1 Conceptual Issues and Their Computational Analogues

We assumed "a priori" that the state space was finite. As noted above without further restrictions models with asymmetric information will generally have to have policies that depend on all past history in order to ensure the equilibrium is perfect; to ensure that if my competitors' condition on a particular public history of play, I can do no better than to condition on that same public history.[19] We noted a number of different rationales for restricting the history that agents can condition their play on. Perhaps most telling among them is that agents have limited memory and/or ability to make computations, and as a result do not have the capacity to either store too long a history, and/or to form differing perceptions of expected discounted values for information sets that detail too much history. Additionally it might be the case that the finite state space is rich enough to adequately approximate an infinite state space, so there is very little to gain by incorporating more detailed information sets. The example investigates this latter possibility in the context of the model just described [see also the discussion in Ifrach and Weintraub (2014)].

To determine whether or not we have an adequate approximation to a perfect equilibrium, we first need to compute a perfect equilibrium to which we can compare to. Fershtman and Pakes (2012) prove that one way to ensure that there is a perfect equilibrium with a finite state space is to assume that there is full revelation of information every T periods. The policies generated by this equilibrium condition only on: the information revealed in the revelation period, the public information that has accumulated since revelation, and the current private information.

[19]To see this in the current model note that whether firm 1 bids in during a particular period depends on whether it thinks firm 2 will bid in, since if firm 2 does not bid in the price firm 1 receives for its electricity will be higher. Firm 1's perception of whether firm 2 bids in will depend on the last time firm 2 did not bid in, as this is the only time it could have done maintenance. If we go back to the period of when firm 2 did not bid in, its decision at that time depended on whether it thought firm 1 would bid in, which depended on the time before that at which firm 1 bid in, and so on. This recursion on the importance of past information set can go on indefinitely.

Table 3.2 Periodic full revelation with different T

	$T = 3$	$T = 4$	$T = 5$	$T = 6$
Summary statistics				
Consumer surplus ($\times 10^{-3}$) 58,000+	550	572	581	580
Profit B ($\times 10^{-3}$)	393	389	384	383
Profit S ($\times 10^{-3}$)	334	324	322	324
Maintenance cost B ($\times 10^{-3}$)	25.9	21.6	20.2	19.4
Maintenance cost S ($\times 10^{-3}$)	12.1	11.8	11.8	11.8
Production cost B ($\times 10^{-3}$)	230.2	235.3	235.1	234.3
Production cost S ($\times 10^{-3}$)	230.4	226.9	228.1	229.2

They then compute policies that are functions of (1) the same variables but with revelation occurring at different T, and (2) coarser partitions of the information set than that used to form perfect equilibrium policies. The policies from (1) are perfect equilibrium policies, just different equilibria for each different T. The policies from the calculations in (2) are not perfect "equilibrium" policies, but do satisfy the conditions in Eqs. (3.2) and (3.3) for the restricted information sets. Fershtman and Pakes (2012) then compare statistics of interest that result from simulations using the different policies.

Table 3.2 presents summary statistics from calculating equilibria assuming different T. It is clear from the table that, at least in this problem, the policies from the $T = 5$ equilibria generate results which are very similar to those from $T = 6$, but the polices from the $T = 3$ equilibria, or even those from $T = 4$, generate results which do not adequately approximate the results from the $T = 6$ equilibria. This suggests that conditioning on a sufficiently long history, in our case $T = 5$, will generate policies that are sufficiently close approximations to the results from policies based on yet longer histories, though I have not made any attempt to formally prove this supposition.

Tables 3.3 and 3.4 present results from two coarser partitions of information sets than the full information set for the $T = 5$ model. In the columns labeled "Finite History T" the results are from a model where there is no revelation of private information and the competitors only keep track of the information publicly revealed in the last $T = 5$ periods. In the columns labeled "Finite History G" the results are from a model where the agents only keep track of the last time each of its competitors' generators was not bid in.

Table 3.3 gives the size of the recurrent class and compute times for the three models. Recall that the only points that are visited repeatedly in the algorithm are in the recurrent class. Table 3.3 provides both the number of such points and the compute time per 100 million iterations of the algorithm. There are two findings to keep in mind; the size of the recurrent class depends on the fineness of the partition of the information sets, and the compute time per million iterations varies directly with the size of the recurrent class. More precisely compute times per a hundred million iterations are increasing and concave in the number of points in the

Table 3.3 Cardinality of $\mathscr{R}$ and compute times

	Finite history		Equilibrium
	G	T	(full revelation)
Cardinality of recurrent class			
1. Firm B ($\times 10^6$)	5650	38,202	67,258
2. Firm S ($\times 10^6$)	5519	47,304	137,489
Compute times per 100 million iterations (h; includes test)			
3. Hours	3:04	11:08	17:14
Hours (100 million)/size of recurrent class (in thousands)			
4. = 3./(1.+2.)	0.26	0.130	0.083

Table 3.4 Three asymmetric information models

	Finite history		Equilibrium
	G	T	(full revelation)
Summary statistics			
Consumer surplus ($\times 10^{-3}$) 58,000+	270	580	581.5
Profit B ($\times 10^{-3}$)	414	384.7	384.5
Profit S ($\times 10^{-3}$)	439	323.5	322.8
Maintenance cost B ($\times 10^{-3}$)	28.5	20.0	20.2
Maintenance cost S ($\times 10^{-3}$)	18.0	11.7	11.8
Production cost B ($\times 10^{-3}$)	226.8	235.5	235.1
Production cost S ($\times 10^{-3}$)	254.6	228.4	228.1

recurrent class. A similar result was found in Pakes and McGuire (2001) who used an analogous algorithm to compute a sequence of symmetric information equilibria with increasing market sizes and hence increases in the cardinality of the underlying state space.[20] The relationship between compute times and the size of the recurrent class is largely the result of the time it takes to search for the data in memory associated with a new location; a point I come back to below.

Table 3.4 compares results of interest generated by policies that are a function of coarser partitions of the information then the information set which generates policies which are "perfect" on the recurrent class. It shows that the partitioning implicit in Finite History T is rich enough to give us an accurate picture of the implications of equilibrium play, while that in Finite History G is not. That is, we do not seem to need the partition implicit in the full information set, but we do need a partition of that information set that is "rich enough" to provide an adequate approximation to equilibrium play. Of course the conditioning variables that generate such an approximation are likely to vary from problem to problem.

[20]For example the maximum number of firms active on recurrent points, an endogenous variable which increased with market size, varied from five to ten in those calculations.

The results in Tables 3.3 and 3.4 are of both analytic and behavioral interest. They suggest three conjectures: (1) the monotonicity of the compute times in the size of the recurrent class, (2) the coarser the partition of the state space the smaller the recurrent class, and (3) coarser partitions that are sufficiently rich can provide adequate approximations to optimal policies. Taken together these well may help explain why the decision makers themselves might partition the information available to them in less detailed ways then they could. It also further emphasizes the question of whether we can investigate the issue of the appropriate conditioning set empirically, a task not attempted in Fershtman and Pakes (2012).

Computational Methods and Burdens

There are many computational issues left to be explored. Two that seemed important determinants of compute time in the work I have been involved in are (1) the way information is stored, and (2) the relationship between initial conditions and the computational burden of the algorithm. For storage we have found that storing the public information with a tree structure and the private information with a hash table conditional on public information worked better than using only one or the other of these two possibilities.

Not surprisingly, we have found that if one starts with sufficiently high values for the initial conditions of the algorithm, that is for the components of $\mathscr{W}$ and Π, the algorithm's iterations will explore almost all possible sample paths.[21] As a result the equilibrium which it eventually generates will typically satisfy the "boundary consistency" condition given in Sect. 3.3.3. This tends to ensure we are not supporting the equilibrium by misperceptions of values at boundary points; though whether actual equilibria are supported by such misperceptions is ultimately an empirical question. On the other hand the higher the initial conditions the longer the compute times before the test in Sect. 3.3.1 is likely to be satisfied. Further, in any given application we can now test whether using smaller starting values results in an equilibrium which is boundary consistent by using the testing procedure discussed in Asker et al. (2014). That is, we can determine whether any set of policies are supported by unrealistic beliefs on outcomes outside of the recurrent class.

This suggest a number of possibilities for reducing the computational burden. For example it may be efficient to use functional form approximations, at least for points outside of the recurrent class, as has been explored in the operations research literature (see Sutton and Barto 1998). Alternatively, the results above indicate that it might be helpful to start out by computing policies that satisfy the test in Sect. 3.3.1 with a coarse partition of the information set. A second step would use those policies as starting values for computing policies for the full information set, or a finer partition of this set.

[21] For the results we discuss below we set $\pi_i^{E,k=0}(m_i, J_i) = \pi_i(m_i, m_{-i} = 0, d, \omega_i)$, and $W^{k=0}(\eta_i, m_i|J_i) = \pi_i(m_i, m_{-i} = 0, d, \omega_i + \eta_i(m_i))/(1-\beta)$. They are based on 500 million iterations and generated an $\mathscr{L}^2(\mathscr{P}(\mathscr{R}))$ norm, i.e., a weighted R^2, of over 0.99995. The $\mathscr{L}^2(\mathscr{P}(\mathscr{R}))$ norm was larger than 0.99 at about 200 and flattened out to the minimum between 250 and 350 million, depending on the run.

3.5 Conclusion

This paper is meant as a contribution to the development of empirical models for the dynamics of market interactions. It argues for a more realistic framework for that analysis, a framework that does not require agents to either acquire and retain excessive amounts of information, or to compute or learn excessively complicated strategies. What we do require is that agents do not make consistent errors conditional on the information they use to compute their policies. The hope is that this weakening of the traditional restrictions of equilibrium play enables both a better approximation to agents behavior and an analytically more convenient framework for the analyst to use in analyzing that behavior empirically.

References

Ackerberg D, Benkard L, Berry S, Pakes A (2007) Econometric tools for analyzing market outcomes. In: Heckman J, Leamer E (eds) The handbook of econometrics. North-Holland, Amsterdam, p 41714276

Arcidiano P, Miller R (2011) Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. Econometrica 7(6):1823–1868

Asker J, Fershtman C, Jeon J, Pakes A (2014) The competitive effects of information sharing. Harvard University (in process, Harvard University)

Bajari, P, Benkard L, Levin J (2007) Estimating dynamic models of imperfect competition. Econometrica 75(5):1331–1370

Benkard L (2004) A dynamic analysis of the market for wide-bodied commercial aircraft. Rev Econ Stud 71:581–611

Benkard L, Weintraub G, Van Roy B (2008) Markov perfect industry dynamics with many firms. Econometrica 79(6):1375–1411

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. Econometrica 63(4):841–890

Besanko D, Doraszelski U, Kryukov Y, Satterthwaite M (2010) Learning by doing, organizational forgetting and industry dynamics. Econometrica 78(2):453–508

Besanko D, Doraszelski U, Kryukov Y (2014) The economics of predation: what drives pricing when there is learning-by-doing? Am Econ Rev 104(3):868897

Collard-Wexler A (2013) Demand fluctuations in the ready-mix concrete industry. Econometrica 81(3):1003–1037

Covert T (2014) Experiential and social learning in firms: the case of hydraulic fracturing in the Bakken Shale. Working paper. University of Chicago, Graduate School of Business

Crawford G, Shum M (2005) Uncertainty and learning in pharmaceutical demand. Econometrica 73(4):1137–1173

Crawford G, Yurukoglu A (2013) The welfare effects of bundling in multichannel television markets. Am Econ Rev 102(2):643–685

Doraszelski U, Judd K (2011) Avoiding the curse of dimensionality in dynamic stochastic games. Quant Econ 3(1):53–93

Doraszelski U, Lewis G, Pakes A (2014) Just starting out: learning and price competition in a new market. Working paper. Harvard University

Einav L, Jenkins M, Levin J (2012) Contract pricing in consumer credit markets. Econometrica 80(4):1387–1432

Ericson R, Pakes A (1995) Markov perfect industry dynamics: a framework for empirical work. Rev Econ Stud 62(1):53–82

Fershtman C, Pakes A (2000) A dynamic game with collusion and price wars. Rand J Econ 31(2):207–236

Fershtman C, Pakes A (2012) Dynamic games with asymmetric information: a framework for applied work. Q J Econ 127(4):1611–1661

Freedman D (1971) Markov chains (Holden-Day series in probability and statistics). Holden-Day, San Francisco

Fudenberg D, Levine D (1983) Subgame perfect equilibrium of finite and infinite horizon games. J Econ Theory 31(2):227–256

Hansen, L, Singleton K (1982) Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50(5):1269–1286

Ifrach B, Weintraub G (2014) A framework for dynamic oligopoly in concentrated industries. Working paper. Columbia University

Judd K (1998) Numerical methods in economics. MIT Press, Cambridge

Kalouptsidi M (2014) Time to build and fluctuations in bulk shipping. Am Econ Rev 104(2): 564–608

Maskin E, Tirole J (1988a) A theory of dynamic oligopoly, I: overview and quantity competition with large fixed costs. Econometrica 56(3):549–569

Maskin E, Tirole J (1988b) A theory of dynamic oligopoly, II: price competition, kinked demand curves, and edgeworth cycles. Econometrica 56(3):571–599

McFadden D (1989) A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57(5):995–1026

Mermelstein B, Nocke V, Satterthwaite M, Whinston M (2014) Internal versus external growth in industries with scale economies: a computational model of optimal merger policy, NBER Working Papers 20051. National Bureau of Economic Research, Inc.

Olley S, Pakes A (1996) The dynamics of productivity in the telecommunications equipment industry. Econometrica 64(6):1263–1298

Pakes A (2010) Alternative models for moment inequalities. Econometrica 78(6):1783–1822

Pakes A, McGuire P (1994) Computing Markov perfect Nash equilibrium: numerical implications of a dynamic differentiated product model. Rand J Econ 259(4):555–589

Pakes A, McGuire P (2001) Stochastic algorithms, symmetric Markov perfect equilibria, and the ‘curse’ of dimensionality. Econometrica 69(5):1261–1281

Pakes A, Pollard D (1989) Simulation and asymptotics of optimization estimators. Econometrica 57(5):1027–1057

Pakes A, Porter J (2014) Moment inequalities for multinomial choice with fixed effects. Working paper. Harvard University

Pakes A, Ostrovsky M, Berry S (2007) Simple estimators for the parameters of discrete dynamic games (with entry-exit examples). Rand J Econ 38(2):373–399

Pakes A, Porter J, Ho K, Ishii J (2015) Moment inequalities and their application. Econometrica 83(1):315–334

Powell J (1994) Estimation of Semiparametric Models, in *Handbook of Econometrics, Vol. 4*, Engle, Robert and Daniel McFadden, eds. Amsterdam: North Holland

Robbins H, Monro S (1951) A stochastic approximation method. Ann Math Stat 22(3):400–407

Rust J (1994) Structural estimation of Markov decision processes. In: Engle R, McFadden D (eds) Handbook of econometrics, vol IV. Elsevier Science, Amsterdam

Sutton R, Barto A (1998) Reinforcement learning: an introduction. MIT Press, Cambridge

White C, Scherer W (1994) Finite-memory suboptimal design for partially observed Markov decision processes. Oper Res 42(3):439–455

Wollmann T (2015) Trucks without bailouts: equilibrium product characteristics for commercial vehicles. Working paper. Harvard University

Chapter 4
Stochastic Games with Signals

Eilon Solan and Bruno Ziliotto

Abstract We survey old and new results concerning stochastic games with signals and finitely many states, actions, and signals. We state Mertens' conjectures regarding the existence of the asymptotic value and its characterization, and present Ziliotto's (Ann Probab, 2013, to appear) counter, example for these conjectures.

Keywords Stochastic games • Signals • Mertens' conjectures • Asymptotic value • Uniform value • Counter example

MSC Classification: 91A15, 91A05

4.1 Introduction

Stochastic games is a model for dynamic interactions in which the state of nature evolves in a way that depends on the actions of the players. The model was first introduced by Shapley (1953), who proved that two-player zero-sum discounted games have a value and both players have optimal stationary strategies. Bewley and Kohlberg (1976) proved that the limit of the discounted value, as the discount factor goes to 0, exists, and is equal to the limit of the value of the n-stage game, as n goes to infinity. This limit is called the asymptotic value of the game. Mertens and Neyman (1981) further showed that for every $\epsilon > 0$ Player 1 (resp. Player 2) has a (history dependent) strategy, which guarantees that the payoff in any sufficiently

E. Solan
School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997800, Israel
e-mail: eilons@post.tau.ac.il

B. Ziliotto (✉)
CEREMADE, Université Paris Dauphine, PSL, Place du Maréchal de Lattre de Tassigny, 75016 Paris, France
e-mail: ziliotto@math.cnrs.fr; brunoziliotto01@gmail.com

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_4

long game, as well as in any discounted game with discount factor sufficiently close to 0, is at least (resp. at most) the asymptotic value minus ϵ (resp. plus ϵ). Such a strategy is called uniform ϵ-optimal.

Mertens et al. (1994) presented a general model of stochastic games with signals, in which the players neither observe the state nor the actions of the other player, but rather observe at every stage a signal that depends on the current state as well as on the pair of actions chosen by the players. Mertens (1986) made the following two conjectures concerning stochastic games with signals and finitely many states, actions, and signals:

- In every stochastic game with signals, the limit of the discounted value, as the discount factor goes to 0, exists, and is equal to the limit of the value of the n-stage game, as n goes to infinity. In other words, the asymptotic value exists.
- If the signal that Player 2 receives is included in the signal that Player 1 receives, then the asymptotic value is equal to the max-min value of the game, which is the maximal quantity that Player 1 can uniformly guarantee in every sufficiently long finite game as well as in every discounted game, provided the discount factor is sufficiently close to 0.

These two conjectures proved to be influential to game theory, and in the attempt to prove them various new tools have been introduced to the field. The conjectures have been shown to hold in quite a few classes of stochastic games with signals (see, e.g., Gensbittel et al. 2014; Neyman 2008; Renault 2006, 2012; Rosenberg 2000; Rosenberg and Vieille 2000; Rosenberg et al. 2002, 2003, 2004; Sorin 1984, 1985; Venel 2014).

Recently Ziliotto (2013) provided an example in which the limit of the discounted value, as the discount factor goes to 0, as well as the limit of the value of the n-stage game, as n goes to infinity, do not exist. In particular, Mertens' conjectures have been refuted.

In this paper we survey the topic of stochastic games with signals and finitely many states, actions, and signals, with an emphasis on the asymptotic value, and present Ziliotto's (2013) example.

4.2 Zero-Sum Standard Stochastic Games

4.2.1 The Model

A two-player zero-sum standard *stochastic game* is described by:

- The set of players $I = \{1, 2\}$.
- A finite state space S.
- For each player $i \in I$ and every state $s \in S$, a finite set of actions $A^i(s)$ that are available to player i at state s. The set of action pairs available at state s is $A(s) := A^1(s) \times A^2(s)$, and the set of all pairs (state, action pair) is $\Lambda := \{(s, a) \colon a \in A(s)\}$.

- A payoff function $u\colon \Lambda \to \mathbb{R}$.
- A transition function $q\colon \Lambda \to \Delta(S)$, where $\Delta(X)$ is the set of probability measures over X for every finite set X.

Given an initial state $s_1 \in S$, the game $\Gamma(s_1)$ proceeds as follows. At each stage $m \geq 1$, each player chooses an action $a^i_m \in A^i(s_m)$, and a new state s_{m+1} is chosen according to the probability measure $q(s_m, a_m)$, where $a_m := (a^1_m, a^2_m)$.

The *history up to stage m* is the sequence $(s_1, a_1, s_2, a_2, \ldots, s_m)$, and the set of all *histories of length m* is $H_m := \Lambda^{m-1} \times S$.

In this section we assume perfect monitoring; that is, at the end of each state m the players observe the new state s_{m+1} and the pair of actions that were just played a_m. Throughout the paper we assume that players have perfect recall; that is, they do not forget information that they learn along the game. Consequently, a strategy σ^i for player i assigns a probability measure over the set of available actions to each finite history. That is, it is a function $\sigma^i\colon \cup_{m\geq 1} H_m \to \cup_{s\in S}\Delta(A^i(s))$ such that for every $m \in \mathbb{N}$ and every finite history $h_m = (s_1, a_1, s_2, a_2, \ldots, s_m) \in H_m$ we have $\sigma^i(h_m) \in \Delta(A^i(s_m))$. The set of strategies for player i is denoted by Σ^i and the set of strategy pairs is $\Sigma := \Sigma^1 \times \Sigma^2$.

An initial state $s_1 \in S$ and a pair of strategies $\sigma \in \Sigma$ induce a probability measure $\mathbb{P}_{s_1,\sigma}$ over the set of all *plays* $H_\infty := \Lambda^{\mathbb{N}}$. The corresponding expectation operator is $\mathbb{E}_{s_1,\sigma}$. For every discount factor $\lambda \in (0, 1]$, the λ*-discounted payoff* under the strategy pair σ at the initial state s_1 is

$$\gamma_\lambda(s_1, \sigma) := \mathbb{E}_{s_1,\sigma}\left(\sum_{m\geq 1} \lambda(1-\lambda)^{m-1} u(s_m, a_m)\right).$$

For every positive integer $n \in \mathbb{N} = \{1, 2, \ldots\}$, the *n-stage payoff* under the strategy pair σ at the initial state s_1 is

$$\gamma_n(s_1, \sigma) := \mathbb{E}_{s_1,\sigma}\left(\frac{1}{n}\sum_{m=1}^{n} u(s_m, a_m)\right).$$

The game $\Gamma_\lambda(s_1)$ is the normal form game $(I, \Sigma^1, \Sigma^2, \gamma_\lambda(s_1, .))$, and the n-stage game $\Gamma_n(s_1)$ is the normal form game $(I, \Sigma^1, \Sigma^2, \gamma_n(s_1, .))$.

4.2.2 The Value

Definition 1. Let $\lambda \in (0, 1]$ be a discount factor, let $n \in \mathbb{N}$ be a positive integer, and let $s_1 \in S$ be the initial state. The real number $v_\lambda(s_1)$ is the *value* of the λ-discounted game $\Gamma_\lambda(s_1)$ if

$$v_\lambda(s_1) = \max_{\sigma^1 \in \Sigma^1} \min_{\sigma^2 \in \Sigma^2} \gamma_\lambda(s_1, \sigma^1, \sigma^2) \tag{4.1}$$

$$= \min_{\sigma^2 \in \Sigma^2} \max_{\sigma^1 \in \Sigma^1} \gamma_\lambda(s_1, \sigma^1, \sigma^2). \tag{4.2}$$

The real number $v_n(s_1)$ is the *value* of the n-stage game $\Gamma_n(s_1)$ if

$$v_n(s_1) = \max_{\sigma^1 \in \Sigma^1} \min_{\sigma^2 \in \Sigma^2} \gamma_n(s_1, \sigma^1, \sigma^2) \tag{4.3}$$

$$= \min_{\sigma^2 \in \Sigma^2} \max_{\sigma^1 \in \Sigma^1} \gamma_n(s_1, \sigma^1, \sigma^2). \tag{4.4}$$

When the initial state is chosen according to a probability distribution $p \in \Delta(S)$, the discounted (resp. n-stage) value is denoted by $v_\lambda(p)$ (resp. $v_n(p)$). In this case, $v_\lambda(p) = \sum_{s \in S} p(s) v_\lambda(s)$ and $v_n(p) = \sum_{s \in S} p(s) v_n(s)$, provided the discounted and n-stage values exist for all initial states.

A strategy $\sigma^1_* \in \Sigma^1$ (resp. $\sigma^2_* \in \Sigma^2$) that attains the maximum (resp. minimum) in Eq. (4.1) [resp. Eq. (4.2)] is called a *λ-discounted optimal strategy*. Similarly, a strategy $\sigma^1_* \in \Sigma^1$ (resp. $\sigma^2_* \in \Sigma^2$) that attains the maximum (resp. minimum) in Eq. (4.3) [resp. Eq. (4.4)] is called an *n-stage optimal strategy*.

A strategy is *stationary* if $\sigma^i(h_m)$ is a function of s_m, for every $m \in \mathbb{N}$ and every finite history $h_m = (s_1, a_1, s_2, a_2, \ldots, s_m) \in H_m$. A strategy is *Markovian* if $\sigma^i(h_m)$ is a function of s_m and m, for every $m \in \mathbb{N}$ and every finite history $h_m = (s_1, a_1, s_2, a_2, \ldots, s_m) \in H_m$. The following two results assert the existence of the value and of stationary (resp. Markovian) optimal strategies in the discounted (resp. n-stage) game.

Theorem 1 (Shapley 1953). *In every standard stochastic game, for every initial state, the λ-discounted value exists. Moreover, both players have stationary strategies that are optimal for all initial states.*

Theorem 2 (Neumann 1928). *In every standard stochastic game, for every initial state, the n-stage value exists. Moreover, both players have Markovian strategies that are optimal for all initial states.*

4.2.3 Zero-Sum Standard Stochastic Games with Long Duration

Considerable effort has been invested on studying properties of stochastic games with long duration, and trying to understand how the value and optimal strategies evolve as the duration goes to infinity. In the discounted game this corresponds to the case where λ converges to 0, and in the n-stage game to the case where n goes to infinity.

In other words, we ask whether there is a quantity w that players can guarantee in every discounted game $\Gamma_\lambda(s_1)$, provided λ is sufficiently close to 0, and in every n-stage game $\Gamma_n(s_1)$, provided n is sufficiently large.

Two approaches can be singled out, the asymptotic approach and the uniform approach. The asymptotic approach assumes that players know the discount factor λ (resp. the length of the game n) and that the discount factor is close to 0 (resp. the length is very large). Consequently, this approach is interested in whether the two limits $\lim_{\lambda \to 0} v_\lambda(s_1)$ and $\lim_{n \to \infty} v_n(s_1)$ exist and are equal.

The uniform approach assumes that the discount factor is close to 0 (resp. the length of the game is very large), yet it does not assume that players know the discount factor λ (resp. the length of the game n). Consequently, this approach is interested in the existence of a strategy that simultaneously guarantees (at least approximately) a payoff greater than $\lim_{\lambda \to 0} v_\lambda(s_1)$ and $\lim_{n \to \infty} v_n(s_1)$ in all games $\Gamma_\lambda(s_1)$ and $\Gamma_n(s_1)$, provided that λ is sufficiently close to 0 and n is sufficiently large.

Definition 2. A stochastic game Γ has an *asymptotic value* if (v_n) and (v_λ) converge (pointwise) to the same limit.

Bewley and Kohlberg (1976) proved that for every initial state s_1, the function $\lambda \to v_\lambda(s_1)$ is a semi-algebraic function (thus continuous at 0), and deduced the following theorem.

Theorem 3 (Bewley and Kohlberg 1976). *Any standard stochastic game has an asymptotic value.*

Definition 3. Let $s_1 \in S$ be a state and let $\alpha \in \mathbb{R}$ be a real number. Player 1 (resp. Player 2) can *uniformly guarantee α at the initial state s_1* if for every $\epsilon > 0$ there exist a strategy $\sigma^1_* \in \Sigma^1$ (resp. $\sigma^2_* \in \Sigma^2$) and a positive integer $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ and every strategy $\sigma^2 \in \Sigma^2$ (resp. $\sigma^1 \in \Sigma^1$),

$$\gamma_n(s_1, \sigma^1_*, \sigma^2) \geq \alpha - \epsilon \quad (\text{resp.} \quad \gamma_n(s_1, \sigma^1, \sigma^2_*) \leq \alpha + \epsilon). \tag{4.5}$$

The real number α is the *uniform value at the initial state s_1* if both players can uniformly guarantee α at s_1. A strategy σ^1_* (resp. σ^2_*) that satisfies (4.5) is called *uniform ϵ-optimal strategy*.

The uniform value at the initial state s_1, when it exists, is denoted by $v_\infty(s_1)$. If a stochastic game has a uniform value at every initial state, then it has an asymptotic value, and both (v_n) and (v_λ) converge pointwise to v_∞ (see Sorin 2002, Chap. 2).

Theorem 4 (Mertens and Neyman 1981). *Any standard stochastic game has a uniform value.*

This result extends to a game with random duration, in which the duration is long in expectation and is independent of the play (see Neyman and Sorin 2010).

Player 1 \ Player 2	L	R
T	1^*	0^*
B	ω	ω

Player 1 \ Player 2	L	R
T	1	0
B	0	1

Fig. 4.1 Transition function (*left*) and payoff function (*right*) in state ω

4.2.4 An Example: The "Big Match"

Consider the following stochastic game, known as the "Big Match", which was introduced by Gillette (1957). The state space is $S = \{\omega, 1^*, 0^*\}$, where 1^* (resp. 0^*) is an absorbing state with payoff 1 (resp. 0): in each of these states, each of the players has a single action, say $A^1(1^*) = \{T\}$ and $A^2(1^*) = \{L\}$ (resp. $A^1(0^*) = \{T\}$ and $A^2(0^*) = \{L\}$), and once the play moves to state 1^* (resp. 0^*), it remains there: $q(1^* \mid 1^*, T, L) = 1$ (resp. $q(0^* \mid 0^*, T, L) = 1$). The action sets for the players in state ω are $A^1(\omega) = \{T, B\}$ and $A^2(\omega) = \{L, R\}$. The payoff and transition functions in this state are described in Fig. 4.1.

For example, if the action pair (T, L) is played at state ω, then the stage payoff is 1 and the play moves to state 1^*, where it stays forever. Using Shapley (1953) one can show that $v_\lambda(\omega) = \frac{1}{2}$ for every $\lambda \in (0, 1]$, and by induction one can show that $v_n(\omega) = \frac{1}{2}$ for every $n \in \mathbb{N}$. In particular, the game has an asymptotic value, which is $\frac{1}{2}$. The stationary strategy $[\frac{1}{2}(L), \frac{1}{2}(R)]$ is an optimal strategy for Player 2 in $\Gamma_\lambda(\omega)$ and $\Gamma_n(\omega)$, for every $\lambda \in (0, 1]$ and every $n \in \mathbb{N}$, and in particular it is a uniform 0-optimal strategy at the initial state ω.

The stationary strategy $[\frac{\lambda}{1+\lambda}(T), \frac{1}{1+\lambda}(B)]$ is an optimal strategy for Player 1 in $\Gamma_\lambda(\omega)$. The time-dependent strategy that plays at stage t the mixed action $[\frac{1}{n-t+2}(T), \frac{n-t+1}{n-t+2}(B)]$ is optimal for Player 2 in $\Gamma_n(\omega)$.

Given $\epsilon < 1/2$, constructing a uniform ϵ-optimal strategy for Player 1 is quite tricky. One can show that Player 1 has no stationary or Markovian strategy that is uniform ϵ-optimal at the initial state ω, nor does he have a uniform ϵ-optimal strategy that can be implemented by a finite automaton. It follows from Blackwell and Ferguson (1968) that given a positive integer M, the following history-dependent strategy for Player 1 is uniform $\frac{1}{2(2M+1)}$-optimal at ω: at stage m, if the play is at state ω, play T with probability $\frac{1}{(M+R_m-L_m)^2}$, where R_m (resp. L_m) is the number of stages up to stage m in which Player 2 played the action R (resp. L). Thus, Player 1 adapts the probability in which he plays T to Player 2's past behavior: as the difference between the number of times that Player 2 played R and the number of times that he played L increases, Player 1's total payoff increases as well, and therefore he lowers the probability to play T and end the play at an absorbing state.

4.3 Zero-Sum Stochastic Games with Signals

So far we assumed that players observe both the current state and past choices of the other player. In many situations, this assumption is unrealistic. For instance, if the state represents a resource stock (like the amount of oil in an oil field), the quantity left, which represents the state, can be evaluated, but is not exactly known. Similarly, various decisions of firms that affect the market price are often not observed by other firms. In this section we extend the model of stochastic games with perfect monitoring to the case in which players do not perfectly observe the state or the actions (see Mertens et al. 1994).

4.3.1 The Model

A *stochastic game with signals* is similar to a standard stochastic game as defined in Sect. 4.2.1 with the following changes:

- There are two finite sets of signals, C for Player 1 and D for Player 2.
- The transition function is a function $q \colon \Lambda \to \Delta(S \times C \times D)$.

At every stage m, each player i chooses an action $a^i_m \in A^i(s_m)$, and a triplet $(s_{m+1}, c_m, d_m) \in S \times C \times D$ is drawn according to the probability measure $q(s_m, a^1_m, a^2_m)$. Player 1 (resp. Player 2) observes the signal c_m (resp. d_m) and the new state is s_{m+1}. We emphasize that the only information that Player 1 (resp. Player 2) has at stage m is the initial state

(or the probability distribution according to which the initial state is chosen), the sequence of past actions that he played, and the sequence of past signals that he received.

A *history* at stage m is a vector $(s_1, a_1, c_1, d_1, s_2, a_2, c_2, d_2, \cdots, s_m)$ and a *play* is a vector in $(\Lambda \times C \times D)^{\mathbb{N}}$. Since players have private information, the *private history* of Player 1 (resp. Player 2) at stage m is $(s_1, a_1, c_1, a_2, c_2, \cdots, a_{m-1}, c_{m-1})$ (resp. $(s_1, a_1, d_1, a_2, d_2, \cdots, a_{m-1}, d_{m-1})$).

Since a player knows the set of actions available for him at every stage of the game, we assume that the private history of a player uniquely identifies his set of actions. For Player 1 this condition translates as follows: for every two histories $(s_1, a_1, c_1, d_1, s_2, a_2, c_2, d_2, \cdots, s_m)$ and $(s'_1, a'_1, c'_1, d'_1, s'_2, a'_2, c'_2, d'_2, \cdots, s'_m)$, if $a^1_t = a'^1_t$ and $c_t = c'_t$ for $1 \le t < m$ then $A_1(s_m) = A_1(s'_m)$. The condition for Player 2 is analogous.

Many models that have been studied in the literature are special cases of stochastic games with signals. These include:

1. Standard stochastic games. These are stochastic games with signals in which the signal contains the new state and the actions that were just played: $C = S \times \{\cup_{s \in S} A(s)\}$ and $c_m = d_m = (s_{m+1}, a^1_m, a^2_m)$.

2. Partially observed Markov decision processes, which are stochastic games with signals that involve only one player: $|I| = 1$.
3. Stochastic games with imperfect monitoring. These are stochastic games with signals in which the signal contains the new state, and possibly additional information: for every $s, s' \in S$, every $a \in A(s)$, every $a' \in A(s')$, every $c \in C$, and every $d \in D$, if $q(c \mid s, a) > 0$ and $q(c \mid s', a') > 0$ then $s = s'$ (the signal of Player 1 uniquely identifies the state), and if $q(d \mid s, a) > 0$ and $q(d \mid s', a') > 0$ then $s = s'$ (the signal of Player 2 uniquely identifies the state).
4. Hidden stochastic games. These are stochastic games with signals in which players receive public signals on the state, and the players observe each other's action: $C = D = \cup_{s \in S} A(s) \times C'$ and $c_m = d_m = (a_m, c'_m)$.
5. Stochastic games played in the dark. These are stochastic games with signals in which the players observe neither the new state nor the action of the opponent: $|C| = |D| = 1$.
6. Repeated games with incomplete information on both sides. These are stochastic games with signals in which the state does not change along the play and each player receives a private signal about the state at the outset of the game and no further information about the state afterwards.

A *strategy* for a player is a function that assigns a probability measure over the set of his available actions to every finite private history of the player.

When the game has perfect monitoring, at each stage m the players know the state sm. When the game does not have perfect monitoring, and the signal that a player receives reveals the other player's action, he can form a belief over the state, which is a probability measure over the set of states S. Consider for example Player 1. At the initial stage his belief over states is the Dirac measure on s_1. If his belief at stage m is $\mu_m \in \Delta(S)$, he played the action a^1_m, Player 2 played the action a^2_m (which he, Player 1, observes), and he observed the signal c_m, then his belief μ_{m+1} at the next stage can be calculated by Bayes rule:

$$\mu_{m+1}(s) = \mathbb{P}(s_{m+1} = s \mid \mu_m, c_m, a^1_m, a^2_m) = \frac{\sum_{s' \in S} \mu_m(s') q(s, c_m \mid s', a^1_m, a^2_m)}{\sum_{s' \in S} q(s, c_m \mid s', a^1_m, a^2_m)},$$

where $q(s, c_m \mid s', a^1_m, a^2_m)$ is the marginal probability of (s, c_m) given (s', a^1_m, a^2_m). When the signals of the two players differ, their belief over the state differs as well, and then each player also has a belief over the belief of the other player, each player has a belief over the belief of the other player on his own belief, and so on. This infinite hierarchy of beliefs that arises naturally in stochastic games with signals explains the challenge that their analysis poses. Note that when the signal does not reveal the action of the other player, the player cannot use Bayes rule to calculate his belief, and in fact the player cannot form a belief over the state, unless he knows the strategy used by the other player.

The concepts of asymptotic value, uniform value, and uniform ϵ-optimal strategies are analogous to the definitions provided above. A natural question is whether stochastic games with signals have an asymptotic value or a uniform value.

By Theorem 1, in the λ-discounted game each player has a stationary strategy that is optimal for all initial states. Consequently, the λ-discounted value of a stochastic game with imperfect monitoring is equal to the λ-discounted value of the same game with perfect monitoring. A similar conclusion holds for the n-stage game, because by Theorem 2, in this game each player has a Markovian strategy that is optimal for all initial states. By Theorem 3 it follows that in stochastic games with imperfect monitoring the asymptotic value exists.

Unfortunately the uniform value may fail to exist in stochastic games with imperfect monitoring. Indeed, consider the "Big Match" and assume that Player 1 observes the state but not the actions of Player 2. Whatever be the signals received by Player 2 about the actions of Player 1, Player 2 can uniformly guarantee $\frac{1}{2}$, but he cannot guarantee any quantity lower than $\frac{1}{2}$. Player 1, on the other hand, can uniformly guarantee 0 but not any positive quantity.

This leads us to the following definition.

Definition 4. Let $s_1 \in S$ and let $\alpha \in \mathbb{R}$. Player 2 can *uniformly defend* α if for every $\epsilon > 0$ and every strategy $\sigma^1 \in \Sigma^1$, there exist a strategy $\sigma^2_* \in \Sigma^2$ and a positive integer $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$,

$$\gamma_n(s_1, \sigma^1, \sigma^2_*) \leq \alpha + \epsilon.$$

The real number α is the *uniform max-min value* at the initial state s_1 if Player 1 can uniformly guarantee α and Player 2 can uniformly defend α.

The result of Mertens and Neyman (1981) generalizes in the following way (see Rosenberg et al. 2003 or Coulomb 2003):

Theorem 5. *In any stochastic game with imperfect monitoring, the uniform max-min value exists for every initial state. Moreover, the uniform max-min value depends only on the signals that Player 1 gets, that is, on the marginal distribution of q over $S \times C$.*

4.3.2 Mertens' Conjectures

Two natural questions that arise are the following. Does the asymptotic value exist in every stochastic game with signals? If it does, can we characterize it?

Mertens (see Mertens 1986, p. 1572 and Mertens et al. 1994, Chap. VIII, p. 378 and 386) stated two conjectures. The first involves the existence of the asymptotic value in any stochastic game with signals.

Conjecture 1. Every zero-sum stochastic game with signals has an asymptotic value.

We say that Player 1 is more informed than Player 2 if the signal of Player 1 contains the signal of Player 2. That is, $C = D \times C'$, and $c_m = (d_m, c'_m)$ for every stage m. The second conjecture of Mertens identifies v_∞.

Conjecture 2. In a zero-sum stochastic game with signals where Player 1 is more informed than Player 2, $\lim_{n\to\infty} v_n$ and $\lim_{\lambda\to 0} v_\lambda$ are equal to the max-min value of the game.

The Mertens conjectures have been proven true in numerous special classes of stochastic games with signals, including standard stochastic games (Bewley and Kohlberg 1976; Mertens and Neyman 1981), stochastic games with imperfect monitoring (Rosenberg et al. 2003; Coulomb 2003), repeated games with incomplete information on both sides (Aumann and Maschler 1995; Mertens and Zamir 1971), and partially observed Markov decision processes (Rosenberg et al. 2002). Other classes of stochastic games in which the conjectures have been proven can be found in Gensbittel et al. (2014), Neyman (2008), Renault (2006, 2012), Rosenberg (2000), Rosenberg and Vieille (2000), Rosenberg et al. (2003, 2004), Sorin (1984, 1985), Venel (2014).

Recall that a hidden stochastic game is a stochastic game in which players receive public signals on the state, and the players observe each other's action. In particular, this is a game in which Player 1 has more information than Player 2. Moreover, in such a game, at every stage both player share the same belief over the state. Ziliotto (2013) provided an example of a hidden stochastic game in which $\lim_{n\to\infty} v_n$ and $\lim_{\lambda\to 0} v_\lambda$ do not exist. This example in particular refutes both of Mertens' conjectures. We provide this example in the next subsection.

4.3.3 A Counterexample to the Mertens' Conjectures

Let $s \in S$ be a state. We say that Player 1 (resp. Player 2) *controls* state s if the transition $q(s, a^1, a^2)$ and the payoff $u(s, a^1, a^2)$ are independent of a^2 for every $a^1 \in A^1(s)$ (independent of a^1 for every $a^2 \in A^2(s)$).

Consider the following hidden stochastic game Γ, with state space $\{1^*, 1^{++}, 1^T, 1^+, 0^*, 0^{++}, 0^+\}$, action sets $\{C, Q\}$ for each player, and signal sets $\{D, D'\}$ for each player. The payoff function does not depend on the actions, and is equal to 1 in states 1^*, 1^{++}, 1^T and 1^+, and to 0 in states 0^*, 0^{++} and 0^+. Player 2 controls states 1^{++}, 1^T and 1^+. Player 1 controls states 0^{++} and 0^+. States 0^* and 1^* are absorbing, and the other states are nonabsorbing. Figure 4.2 describes the transition function.

In Fig. 4.2 we adopt the following notation: an arrow going from state s to state s' with the caption $(a, p, c) \in \{C, Q\} \times [0, 1] \times \{D, D'\}$ indicates that if the player who controls state s plays action a, then with probability p the state moves to state s' and the signal is c. For example, if the state is 1^{++} and Player 2 plays action C,

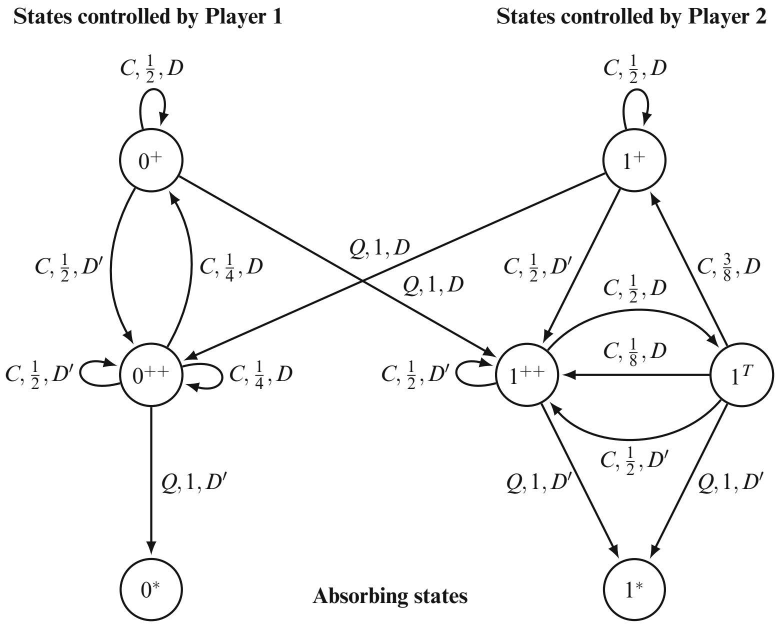

Fig. 4.2 Transitions in the game Γ

then with probability $\frac{1}{2}$ the game moves to state 1^T and the signal is D, and with probability $\frac{1}{2}$ the game stays in state 1^{++} and the signal is D'.

This game is a hidden stochastic game. As mentioned above, in such a game, at every stage the players share the same belief over the states. In particular, we can consider an equivalent auxiliary stochastic game with perfect monitoring but with countably many states; a state in the auxiliary game corresponds to a belief over states in the original game. Since the number of states, actions, and signals is finite, the number of possible beliefs at each stage is finite, so that in the auxiliary game there are countable many states. By Theorems 1 and 2 (generalized to games with countably many states) in the discounted game the players have optimal stationary strategies and in the n-stage game they have optimal Markovian strategies. Any stationary or Markovian strategy in the auxiliary game has an equivalent strategy in the original game, and vice versa, and therefore the original game and the auxiliary game are equivalent in terms of the discounted value, the n-stage value, and optimal strategies.

Call the states $\{1^*, 1^{++}, 1^T, 1^+\}$ *1-states* and the states $\{0^*, 0^{++}, 0^+\}$ *0-states*. In our example, the players know when the play moves from 1-states to 0-states and vice versa. Indeed, the initial state is known, so that the players know whether it is a 1-state or a 0-state. The play moves to a 1-state (resp. 0-state) to a 0-state (resp. 1-state) only when Player 2 (resp. Player 1) plays Q and the signal is D. Consequently the possible beliefs of the players along the play are:

- $[1(0^*)]$ and $[1(1^*)]$: the players know when the play moves to an absorbing state. This is the belief at stage m when the player who controls state s_{m-1} played Q and the signal was D'.
- $[2^{-n}(0^{++}), (1-2^{-n})(0^+)]$ for $n \geq 0$: players believe that with probability 2^{-n}, the state is 0^{++}, and with probability $(1-2^{-n})$, the state is 0^+. For $n = 0$, this is the belief at stage m when (a) the state in stage $m-1$ was a 1-state, Player 2 played Q and the signal was D, or (b) the state in stage $m-1$ was a 0-state, Player 1 played C and the signal was D'. For $n \geq 1$, this is the belief at stage m when the belief at stage $m-1$ was $[2^{-(n-1)}(0^{++}), (1-2^{-(n-1)})(0^+)]$, Player 1 played C and the signal was D.
- $[2^{-2n}(1^{++}), (1-2^{-2n})(1^+)]$ for $n \geq 0$: players believe that with probability 2^{-2n}, the state is 1^{++}, and with probability $(1-2^{-2n})$, the state is 1^+. For $n = 0$, this is the belief at stage m when (a) the state in stage $m-1$ was a 0-state, Player 1 played Q and the signal was D, or (b) the state in stage $m-1$ was a 1-state, Player 2 played C and the signal was D'. For $n \geq 1$, this is the belief at stage m when the belief at stage $m-1$ was $[2^{-2n-2}(1^T), (1-2^{-2n-2})(1^+)]$, Player 2 played C and the signal was D.
- $[2^{-2n}(1^T), (1-2^{-2n})(1^+)]$ for $n \geq 0$: players believe that with probability 2^{-2n}, the state is 1^T, and with probability $(1-2^{-2n})$, the state is 1^+. This is the belief at stage m when the belief at stage $m-1$ was $[2^{-2n}(1^{++}), (1-2^{-2n})(1^+)]$, Player 2 played C and the signal was D.

To simplify notation, we denote these beliefs as follows:

- 0^* is the belief $[1(0^*)]$; 1^* is the belief $[1(1^*)]$.
- 0_n is the belief $[2^{-n}(0^{++}), (1-2^{-n})(0^+)]$.
- 1_{2n} is the belief $[2^{-2n}(1^{++}), (1-2^{-2n})(1^+)]$.
- 1_{2n+1} is the belief $[2^{-2n}(1^T), (1-2^{-2n})(1^+)]$.

Thus, the auxiliary game is a stochastic game with perfect information that is given by

- The set of states is $\{0^*, 0_0, 0_1, 0_2, 0_3, \cdots, 1^*, 1_0, 1_1, 1_2, 1_3, \cdots\}$.
- In all states the players have two actions, $\{C, Q\}$.
- The payoff in states $\{0^*, 0_0, 0_1, 0_2, 0_3, \cdots\}$ is 0; the payoff in states $\{1^*, 1_0, 1_1, 1_2, 1_3, \cdots\}$ is 1.
- States 0^* and 1^* are absorbing. The transition function in states $\{0_n, n \geq 0\}$ is described in Fig. 4.3 and the transition function in states $\{1_n, n \geq 0\}$ is described in Fig. 4.4.

We will show below that the limit $\lim_{\lambda\to 0} v_\lambda(s)$ does not exist for every nonabsorbing state of the auxiliary game. We thus consider now the discounted game.

Since the action of Player 2 (resp. Player 1) in states $\{0_0, 0_1, 0_2, 0_3, \cdots\}$ (resp. $\{1_0, 1_1, 1_2, 1_3, \cdots\}$) affects neither the payoff nor the transitions, and since the players know the current state of the auxiliary game, for the calculation of the value we can assume that in states $\{0_0, 0_1, 0_2, 0_3, \cdots\}$ only Player 1 chooses an

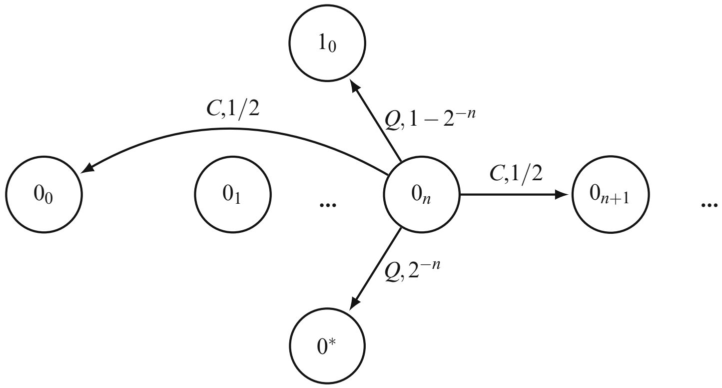

Fig. 4.3 Transitions in the state 0_n

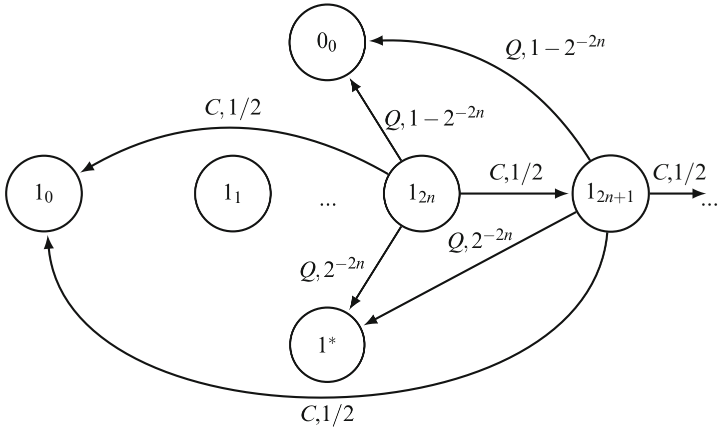

Fig. 4.4 Transitions in the states 1_{2n} and 1_{2n+1}

action, while in states $\{1_0, 1_1, 1_2, 1_3, \cdots\}$ only Player 2 chooses an action. This implies that the decision problems of the two players, namely, the decision problem of Player 1 in states $\{0_0, 0_1, 0_2, 0_3, \cdots\}$ and the decision problem of Player 2 in states $\{1_0, 1_1, 1_2, 1_3, \cdots\}$, can be disentangled into two separate Markov decision problems. It follows that in the discounted game each player has an optimal pure stationary strategy. Denote by σ_n^1 (resp. σ_n^2) the stationary strategy for Player 1 (resp. Player 2) that chooses action C in states $\{0_k, k \neq n\}$ (resp. $\{1_k, k \neq n\}$) and chooses action Q in state 0_n (resp. 1_n).

Assume that the initial state is $p = 0_0$. Since the payoff is 1 in states $\{1_n,\ n \in \mathbb{N}\}$ and 0 in states $\{0_n,\ n \in \mathbb{N}\}$, and since Player 1 maximizes the payoff, Player 1 wants the play to move to state 1_0. If he plays Q in state 0_0, the game is absorbed in state 0^* with probability 1, which is the worst state for him. If he never plays Q, the payoff is 0 forever, which is also an unfavorable outcome for Player 1. If he plays strategy σ_n^1, then the play reaches state 0_n after a random finite number of stages. It is then absorbed in state 0^* with probability 2^{-n} (we will call this probability the *absorbing risk taken by Player 1*), and moves to state 1_0 with probability $(1 - 2^{-n})$.

Inspecting the transition function in Fig. 4.3 reveals that to reach state 0_n from state 0_0, Player 1 needs on average 2^n stages. Thus, Player 1's decision of when to play Q is influenced by two contradicting forces: on the one hand he wants to lower the absorbing risk, which means adopting a strategy σ_n^1 with high n; on the other hand he wants to minimize the time he spends in states $\{0_0, 0_1, 0_2, 0_3, \cdots\}$, which means he should not choose n too high. Since the game is discounted, the discount factor determines the positive integer n such that σ_n^1 is optimal. It turns out that in the λ-discounted game, the optimal strategies for Player 1 is σ_n^1, where 2^{-n} is as close as possible to $\sqrt{2\lambda}$ (the real number $\sqrt{2\lambda}$ corresponds to the optimal absorbing risk if players could choose any absorbing risk in [0, 1], which is not the case).

Player 2 faces the same issue in states $\{1_0, 1_1, 1_2, 1_3, \cdots\}$. However the difference between the transitions in states $\{1_0, 1_1, 1_2, 1_3, \cdots\}$ and in states $\{0_0, 0_1, 0_2, 0_3, \cdots\}$ leads to a slightly different optimal strategy. We claim that the strategy σ_{2n}^2 is strictly better than the strategy σ_{2n+1}^2. Indeed, both strategies exhibit the same absorbing risk, yet the former requires much fewer stages to move to state 1_0 than the latter. In particular, the optimal strategy of Player 2 is σ_{2n}^2 for some positive integer n. In fact, there exists two mappings $\epsilon_1, \epsilon_2 : (0, 1] \to \mathbb{R}$ such that $\lim_{\lambda \to 0} \epsilon_1(\lambda) = \lim_{\lambda \to 0} \epsilon_2(\lambda) = 0$, and if $2^{-2n-1+\epsilon_1(\lambda)} \leq \sqrt{2\lambda} \leq 2^{-2n+1+\epsilon_2(\lambda)}$, then the optimal strategy is σ_{2n}^2.

Set $\lambda_k := 2^{-4k-1}$, so that $\sqrt{2\lambda_k} = 2^{-2k}$. In the game $\Gamma_{\lambda_k}(1_0)$ the optimal strategies of the players are σ_{2k}^1 and σ_{2k}^2, and the play is symmetric: the sequence $(v_{\lambda_k}(1_0))_{k \geq 1}$ converges to $\frac{1}{2}$. Set $\mu_k := 2^{-4k-3}$, so that $\sqrt{2\mu_k} = 2^{-2k-1}$. In this case, Player 1's optimal strategy is σ_{2k+1}^1, yet Player 2's optimal strategy is not σ_{2k+1}^2, because his optimal strategy is taken from the set $\{\sigma_{2n}^2, n \geq 0\}$. Player 2's optimal strategy is either σ_{2k}^2 or σ_{2k+2}^2. But choosing σ_{2k}^2 or σ_{2k+2}^2 instead of σ_{2k+1}^2 leads to a different dynamics of the state. For instance, under the strategy σ_{2k+2}^2, starting from state 1_0 Player 2 waits on average 2^{2k+2} stages before playing Q, instead of 2^{2k+1}. Thus, intuitively, Player 1 has an edge over his opponent in the μ_k-discounted game. Standard computations confirm this intuition, and show that $(v_{\mu_k}(1_0))$ converges to $\frac{5}{9}$. In particular, the limit $\lim_{\lambda \to 0} v_\lambda(1_0)$ does not exist, which implies that the limit $\lim_{\lambda \to 0} v_\lambda(s)$ does not exist for every state in $\{1_n, n \in \mathbb{N}\}$. Similar arguments show that this limit does not exist for every state in $\{0_n, n \geq 0\}$.

4.3.4 Link Between the Convergence of (v_n) and (v_λ)

In the previous example, neither (v_λ) nor (v_n) converge. There is an example of a hidden stochastic game for which there exists an initial belief $p \in \Delta(S)$ such that $(v_\lambda(p))$ converges but $(v_n(p))$ does not, and conversely, an example where $(v_n(p))$ converges and $(v_\lambda(p))$ does not converge. Moreover, there are examples in which both $(v_\lambda(p))$ and $(v_n(p))$ converge, but to different limits (Ziliotto 2015). Nonetheless, Ziliotto (2015) proved the following Tauberian theorem, which is a generalization of the one-player case result of Lehrer and Sorin (1992):

Theorem 6. *Consider the auxiliary game of a hidden stochastic game. The following two statements are equivalent:*

1. *For every initial state* p_1, $\lim_{n\to\infty} v_n(p_1)$ *exists.*
2. *For every initial state* p_1, $\lim_{\lambda\to 0} v_\lambda(p_1)$ *exists.*

Moreover, when these statements hold, we have $\lim_{n\to\infty} v_n(p_1) = \lim_{\lambda\to 0} v_\lambda(p_1)$ *for every initial state* p_1.

This theorem is in fact true in a much wider class of stochastic games with compact state space and actions sets.

4.4 Multiplayer Stochastic Games

A multiplayer stochastic game is similar to a two-player zero-sum stochastic game as defined in Sect. 4.2.1 with the following changes:

- The set of players I is any finite set.
- For every player $i \in I$ and each state s, the set $A^i(s)$ is a finite set of actions available to player i at state s. Denote $A(s) := \times_{i\in I} A^i(s)$ and $\Lambda := \{(s,a) \colon s \in S, a \in A(s)\}$.
- For every player $i \in I$ the payoff function is $u^i : \Lambda \to \mathbb{R}$.

In this case we consider the asymptotic behavior of the set $E_\lambda(s_1)$ of all λ-discounted equilibrium payoffs at the initial state s_1 and of the set $E_n(s_1)$ of all n-stage equilibrium payoffs at the initial state s_1.

4.4.1 Asymptotic Approach

The natural generalization of the result of Bewley and Kohlberg (1976) to the multiplayer case would be the convergence of the set of λ-discounted Nash equilibrium payoffs E_λ when λ goes to 0 and the convergence of the set of n-stage Nash equilibrium payoffs E_n (w.r.t. the Hausdorff distance).

Note that for a repeated game (that is, a stochastic game with one state) that satisfies certain technical conditions, the Folk theorem answers this question, and gives a characterization of the limit set. A Folk theorem for multiplayer stochastic games has been proven by Dutta (1995), under the strong assumption that the dependence of $E_\lambda(s_1)$ in s_1 vanishes as λ goes to 0 (in particular, this excludes the presence of absorbing states with different payoffs).

In the general case, E_λ and E_n may fail to converge, even in the two-player case, as proved by Renault and Ziliotto (2014), who prove in addition that the set of discounted (or n-stage) subgame perfect equilibrium payoffs may fail to converge, but the set of discounted stationary Nash equilibrium payoffs converges.

4.4.2 Uniform Approach

The concept of uniform value can be generalized in the following way to the nonzero-sum case.

Definition 5. A vector $v \in \mathbb{R}^I$ is a uniform equilibrium payoff at the initial state s_1 if for every $\epsilon > 0$ there exist $\lambda_0 \in (0, 1]$, $n_0 \in \mathbb{N}$, and a strategy vector $\sigma \in \Sigma$, such that for every player $i \in I$ and every strategy $\sigma'^i \in \Sigma^i$,

$$\gamma_\lambda^i(s_1, \sigma'^i, \sigma^{-i}) - 2\epsilon \le v^i - \epsilon \le \gamma_\lambda^i(s_1, \sigma) \le v^i + \epsilon, \quad \forall \lambda \in (0, \lambda_0],$$

and

$$\gamma_n^i(s_1, \sigma'^i, \sigma^{-i}) - 2\epsilon \le v^i - \epsilon \le \gamma_n^i(s_1, \sigma) \le v^i + \epsilon, \quad \forall n \ge n_0.$$

Vrieze and Thuijsman (1989) proved the existence of a uniform equilibrium payoff in two-player absorbing games, which are stochastic games with a single nonabsorbing state. Vieille (2000a,b) extended this result to any two-player stochastic game. Flesch et al. (1997) provided an example of a three-player absorbing game in which the uniform equilibrium strategies are periodic. Solan (1999) proved that any three-player absorbing game has a uniform equilibrium in which the players execute a periodic play path, and supplement their play with threats of punishment.

For $N \ge 3$, the existence of a uniform equilibrium payoff in multiplayer stochastic games is still open, and is one of the most important and challenging question in mathematical game theory to date. When players are allowed to use a correlation device, this question was solved positively by Solan and Vieille (2002).

4.4.3 Multiplayer Stochastic Games with Imperfect Monitoring

Like in the zero-sum case (see Sect. 4.3), we assume that players do not observe the actions of the other players, but rather receive signals about them. This model is more general than the one of the previous section, thus (E_λ) may not converge (see Renault and Ziliotto 2014). In the literature, Folk theorems for stochastic games with imperfect monitoring are stated under two kinds of assumptions. The first one is an ergodic assumption on the transition function of the game. The second one is either that players do not use their private information [public equilibrium, see

Hörner et al. (2011), or public signals, see Fudenberg and Yamamoto (2011)], or an assumption on the signaling function (connectedness, as in Fudenberg and Levine 1991).

4.4.4 Stochastic Games with Signals on the State

When the state is imperfectly observed, it does not seem possible to generalize the result of Bewley and Kohlberg (1976). Indeed, Renault and Ziliotto (2014) provided an example of a two-player hidden stochastic game (public signals on the state and perfect observation of the actions), in which E_λ has full dimension for all $\lambda \in (0, 1]$ (which is a standard assumption in the literature under which Folk theorems are usually stated), but the set of discounted correlated equilibrium payoffs, discounted Nash equilibrium payoffs, discounted sequential equilibrium payoffs, discounted stationary equilibrium payoffs, all fail to converge. Under an ergodic assumption, Yamamoto (2015) recently proved a Folk theorem for multiplayer hidden stochastic games.

Acknowledgements Solan acknowledges the support of grant #323/13 of the Israel Science Foundation. Ziliotto acknowledges the support of the ANR Jeudy (ANR-10-BLAN 0112) and the GDR 2932.

References

Aumann RJ, Maschler M (1995) Repeated games with incomplete information. MIT Press, Cambridge
Bewley T, Kohlberg E (1976) The asymptotic theory of stochastic games. Math Oper Res 1(3):197–208
Blackwell D, Ferguson TS (1968) The big match. Ann Math Stat 39(1):159–163
Coulomb J-M (2003) Stochastic games without perfect monitoring. Int J Game Theory 32(1):73–96
Dutta P (1995) A folk theorem for stochastic games. J Econ Theory 66(1):1–32
Flesch J, Thuijsman F, Vrieze K (1997) Cyclic Markov equilibria in stochastic games. Int J Game Theory 26(3):303–314,
Fudenberg D, Levine D (1991) An approximate folk theorem with imperfect private information. J Econ Theory 54(1):26–47
Fudenberg D, Yamamoto Y (2011) The folk theorem for irreducible stochastic games with imperfect public monitoring. J Econ Theory 146(4):1664–1683
Gensbittel F, Oliu-Barton M, Venel X (2014) Existence of the uniform value in repeated games with a more informed controller. J Dyn Games 1:411–445
Gillette D (1957) Stochastic games with zero stop probabilities. Contrib Theory Games 3:179–187
Hörner J, Sugaya T, Takahashi S, Vieille N (2011) Recursive methods in discounted stochastic games: an algorithm for $\delta \to 1$ and a folk theorem. Econometrica 79(4):1277–1318
Lehrer E, Sorin S (1992) A uniform Tauberian theorem in dynamic programming. Math Oper Res 17(2):303–307

Mertens JF (1986) Repeated games. In: Proceedings of the international congress of mathematicians, Berkeley, CA, pp 1528–1577

Mertens JF, Neyman A (1981) Stochastic games. Int J Game Theory 10(2):53–66

Mertens JF, Zamir S (1971) The value of two-person zero-sum repeated games with lack of information on both sides. Int J Game Theory 1(1):39–64

Mertens JF, Sorin S, Zamir S (1994) Repeated games. CORE DP 9420–9422

Neyman A (2008) Existence of optimal strategies in markov games with incomplete information. Int J Game Theory 37(4):581–596

Neyman A, Sorin S (2010) Repeated games with public uncertain duration process. Int J Game Theory 39(1):29–52

Renault J (2006) The value of markov chain games with lack of information on one side. Math Oper Res 31(3):490–512

Renault J (2012) The value of repeated games with an informed controller. Math Oper Res 37(1):154–179

Renault J, Ziliotto B (2014) Hidden stochastic games and limit equilibrium payoffs. arXiv preprint arXiv:1407.3028

Rosenberg D (2000) Zero-sum absorbing games with incomplete information on one side: asymptotic analysis. SIAM J Control Optim 39(1):208–225

Rosenberg D, Vieille N (2000) The maxmin of recursive games with incomplete information on one side. Math Oper Res 25(1):23–35

Rosenberg D, Solan E, Vieille N (2002) Blackwell optimality in markov decision processes with partial observation. Ann Stat 30:1178–1193

Rosenberg D, Solan E, Vieille N (2003) The maxmin value of stochastic games with imperfect monitoring. Int J Game Theory 32(1):133–150

Rosenberg D, Solan E, Vieille N (2004) Stochastic games with a single controller and incomplete information. SIAM J Control Optim 43(1):86–110

Shapley LS (1953) Stochastic games. Proc Natl Acad Sci USA 39(10):1095–1100

Solan E (1999) Three-player absorbing games. Math Oper Res 24(3):669–698

Solan E, Vieille N (2002) Correlated equilibrium in stochastic games. Games Econ Behav 38(2):362–399

Sorin S (1984) Big matchİ with lack of information on one side (part i). Int J Game Theory 13(4):201–255

Sorin S (1985) Big matchİ with lack of information on one side (part ii). Int J Game Theory 14(3):173–204

Sorin S (2002) A first course on zero-sum repeated games. In: Mathématiques et applications, vol 37. Springer, Berlin

Venel X (2014) Commutative stochastic games. Math Oper Res 40(2):403–428

Vieille N (2000a) Two-player stochastic games i: a reduction. Isr J Math 119(1):55–91

Vieille N (2000b) Two-player stochastic games ii: the case of recursive games. Isr J Math 119(1):93–126

von Neumann J (1928) Zur theorie der gesellschaftsspiele. Math Ann 100:295–320

Vrieze OJ, Thuijsman F (1989) On equilibria in repeated games with absorbing states. Int J Game Theory 18(3):293–310

Yamamoto Y (2015) Stochastic games with hidden states, preprint, PIER working paper

Ziliotto B (2013) Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture maxmin= lim v (n). Ann Probab. arXiv preprint arXiv:1305.4778 (to appear)

Ziliotto B (2015) A tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games. arXiv preprint arXiv:1501.06525, Mathematics of Operations Research (to appear)

Chapter 5
Nonlinear and Multiplayer Evolutionary Games

Mark Broom and Jan Rychtář

Abstract Classical evolutionary game theory has typically considered populations within which randomly selected pairs of individuals play games against each other, and the resulting payoff functions are linear. These simple functions have led to a number of pleasing results for the dynamic theory, the static theory of evolutionarily stable strategies, and the relationship between them. We discuss such games, together with a basic introduction to evolutionary game theory, in Sect. 5.1. Realistic populations, however, will generally not have these nice properties, and the payoffs will be nonlinear. In Sect. 5.2 we discuss various ways in which nonlinearity can appear in evolutionary games, including pairwise games with strategy-dependent interaction rates, and playing the field games, where payoffs depend upon the entire population composition, and not on individual games. In Sect. 5.3 we consider multiplayer games, where payoffs are the result of interactions between groups of size greater than two, which again leads to nonlinearity, and a breakdown of some of the classical results of Sect. 5.1. Finally in Sect. 5.4 we summarise and discuss the previous sections.

Keywords ESS • Payoffs • Matrix games • Nonlinearity • Multi-player games

MSC Codes: 91A22, 91A06, 91A80

M. Broom (✉)
Department of Mathematics, City University London, Northampton Square,
London EC1V 0HB, UK
e-mail: mark.broom@city.ac.uk

J. Rychtář
Department of Mathematics and Statistics, The University of North Carolina
at Greensboro, Greensboro, NC 27412, USA
e-mail: rychtar@uncg.edu

© Springer International Publishing Switzerland 2016

F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*,
Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_5

5.1 Introduction

In this paper we consider nonlinear and multiplayer evolutionary games. We start in Sect. 5.1 with an introduction to evolutionary games for those not familiar with them, focusing on matrix games, which are linear in character, and discussing a number of the key results. We then move on to consider the general idea of nonlinear evolutionary games, including some specific types of such games in Sect. 5.2. We believe that these results, and those in the following section, will generally be less familiar to the audience. In Sect. 5.3 we consider multiplayer games. The specific type that we consider, and the most commonly used, is multiplayer matrix games, which can be though of as a special type of the nonlinear games in Sect. 5.2, although we note that multiplayer games in general do not simply reduce to this type. The text in significant part follows a tutorial talk given by MB at the International Society on Dynamic Games Symposium in Amsterdam in July 2014, which in turn followed aspects of the book (Broom and Rychtář 2013).

5.1.1 What is Evolutionary Game Theory?

Evolutionary game theory as we know it today began in the 1960s, in particular with the consideration of the sex-ratio problem (Hamilton 1967), although similar reasoning on this problem goes back much earlier to Dusing (see Edwards 2000) and Fisher 1930. The most influential work on our modern understanding is that of Maynard Smith and collaborators (Maynard Smith et al. 1973; Maynard Smith 1982).

In (non-cooperative) game theory, a game is comprised of three key elements, the *players*, the *strategies* available to be employed by the players, and the *payoffs* to the players, which are functions of the strategies chosen. For an evolutionary game we also need a *population*, and a way for our population to evolve through time, an *evolutionary dynamics*.

A *pure strategy* is a choice of what to play in a given interaction. Supposing that the pure strategies comprise the finite set $\{S_1, S_2, \ldots, S_n\}$, then a mixed strategy is defined as a probability vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$, p_i being the probability that the player will play pure strategy S_i. Thus a pure strategy can be written in this way, e.g. S_i is $(0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 at the ith place, and a mixed strategy can be written as a convex combination of pure strategies,

$$\mathbf{p} = (p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i S_i. \tag{5.1}$$

The set of all mixed strategies can be represented by a simplex in $\mathbb{R}^n$ with vertices at $\{S_1, S_2, \ldots, S_n\}$ (Fig. 5.1). The *Support of* $\mathbf{p}$, $S(\mathbf{p})$, is defined by $S(\mathbf{p}) = \{i : p_i > 0\}$, so that it is the set of pure strategies which have a positive probability of being played by a $\mathbf{p}$-player. The notion of a mixed strategy is naturally extended even to

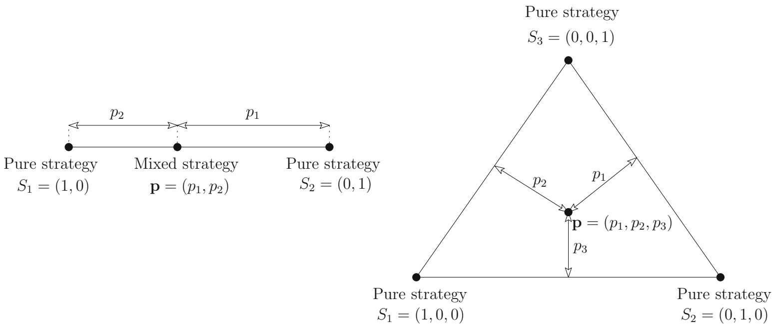

Fig. 5.1 Visualization of pure and mixed strategies for games with two or three strategies

cases where the set of pure strategies is infinite, as in the "war of attrition" game, for example Bishop and Cannings (1978).

Payoffs for a game played by two players with each having a finite number of pure strategies can be represented by two matrices. For example, if player 1 has the strategy set $\mathbf{S} = \{S_1, \ldots, S_n\}$ and player 2 has the strategy set $\mathbf{T} = \{T_1, \ldots, T_m\}$, then the payoffs in this game are written as

$$A = (a_{ij})_{i=1,\ldots,n;j=1,\ldots,m}, B = (b_{ij})_{i=1,\ldots,m;j=1,\ldots,n}, \tag{5.2}$$

where a_{ij} (b_{ji}) is the reward to players 1 (2) after player 1 (2) chooses pure strategy S_i (T_j). We thus have the payoffs written as a pair of $n \times m$ matrices A and B^T, which is known as a bimatrix representation. This is often written as a single matrix whose entries are ordered pairs of values.

Note that here we write the payoffs from the point of view of the player receiving the reward (i.e. the index of their strategy comes first). It is often the case in other works that the index of player 1 is written first.

Often in evolutionary games, the choice of which player is player 1 is arbitrary, and thus the strategies available to the two players are identical. In this case, $n = m$ and (after a possible renumbering) $S_i = T_i$ for all i. Since the ordering of players is arbitrary, if we swap them their payoffs are unchanged, so that $b_{ij} = a_{ij}$, i.e. $A = B$. This means that all payoffs can be written as a single $n \times n$ matrix

$$A = (a_{ij})_{i,j=1,\ldots,n}, \tag{5.3}$$

where in this case, a_{ij} is the payoff to a player playing pure strategy S_i when its opponent plays strategy S_j. Such a game is called a *matrix game*.

Consider a game with payoffs given by a matrix A. If player 1 plays $\mathbf{p}$ and player 2 plays $\mathbf{q}$, then the proportion of games involving the first player playing S_i and the second player playing S_j is simply $p_i q_j$. The expected reward to player 1 is thus given by

$$E[\mathbf{p}, \mathbf{q}] = \sum_{i,j} a_{ij} p_i q_j = \mathbf{p} A \mathbf{q}^{\mathbf{T}}. \tag{5.4}$$

Note that, when comparing payoffs, we can ignore difficult cases involving equalities by assuming our games are *generic* (Samuelson 1997; Broom and Rychtář 2013). In most of the following we will make this assumption.

In the above, we have considered a single game between two individuals. However, evolutionary games consist of populations, and individuals are not (usually) involved in only a single contest. They may play many different contests, against many different opponents, with each contributing a relatively small contribution to the total reward.

We consider a function $\mathscr{E}[\sigma;\Pi]$, the fitness of an individual using a strategy σ in a population represented by Π. The term $\delta_{\mathbf{p}}$ is used to represent a population where the probability of a randomly selected player being a $\mathbf{p}$-player is 1. The term δ_i similarly denotes a population consisting only of individuals playing pure strategy S_i (with probability 1). The term $\sum_i p_i\delta_i$ thus means a population where the proportion of S_i-playing individuals is p_i.

5.1.2 Two Approaches to Game Analysis

5.1.2.1 Dynamic Analysis

In all that follows we assume a very large (effectively infinite) population, with overlapping generations and asexual reproduction, where offspring are direct copies of their parent. The evolution of a population can be modelled using evolutionary dynamics, where the proportion of individuals playing a given strategy changes according to their fitness.

In the following we shall assume a population consisting only of pure strategists. Thus we consider a population represented by $\mathbf{p}^{\mathbf{T}} = \sum_i p_i\delta_i$, i.e. where the frequency of S_i-playing individuals is p_i. We denote the fitness of individuals playing S_i in this population to be $f_i(\mathbf{p})$. The birth rate of individuals in the population is proportional to their fitness.

We assume that the composition of the population changes according to the differential equation

$$\frac{d}{dt}p_i = p_i\Big(f_i\big(\mathbf{p}(t)\big) - \bar{f}(\mathbf{p}(t))\Big). \tag{5.5}$$

This is the *continuous replicator dynamics*, the most commonly used evolutionary dynamics, originating in Taylor and Jonker (1978) (see also Hofbauer and Sigmund 1998). For a derivation see Broom and Rychtář (2013). We also note the existence of the discrete replicator dynamics, the equivalent dynamics for non-overlapping generations (see Bishop and Cannings 1978).

For matrix games the continuous replicator dynamics (5.5) becomes

$$\frac{d}{dt}p_i = p_i\left(\Big(A\big(\mathbf{p}(t)\big)^{\mathbf{T}}\Big)_i - \mathbf{p}(t)A\big(\mathbf{p}(t)\big)^{\mathbf{T}}\right). \tag{5.6}$$

5.1.2.2 Static Analysis

An alternative methodology is to use a static analysis, which does not consider how the population reached a particular point in the strategy space, but assuming that the population is at that point, asks whether other strategies can do better within the population?

Consider a population where the vast majority of individuals play strategy S, while a very small proportion $\varepsilon > 0$ of "mutants" play strategy M. The strategies S and M thus compete within the population $(1-\varepsilon)\delta_S + \varepsilon\delta_M$. A strategy S is *e*volutionarily stable against strategy M if there is $\varepsilon_M > 0$ such that

$$\mathscr{E}[S; (1-\varepsilon)\delta_S + \varepsilon\delta_M] > \mathscr{E}[M; (1-\varepsilon)\delta_S + \varepsilon\delta_M] \tag{5.7}$$

for all $\varepsilon < \varepsilon_M$. S is an *evolutionarily stable strategy (ESS)* if it is *e*volutionarily stable against M for every other strategy $M \neq S$ (Maynard Smith et al. 1973; Maynard Smith 1982).

For matrix games, the linearity of the payoffs gives

$$\mathscr{E}[\mathbf{p}; (1-\varepsilon)\delta_{\mathbf{p}} + \varepsilon\delta_{\mathbf{q}}] = E[\mathbf{p}, (1-\varepsilon)\mathbf{p} + \varepsilon\mathbf{q}] = \tag{5.8}$$

$$\mathbf{p}A((1-\varepsilon)\mathbf{p} + \varepsilon\mathbf{q})^{\mathbf{T}} = (1-\varepsilon)\mathbf{p}A\mathbf{p}^{\mathbf{T}} + \varepsilon\mathbf{p}A\mathbf{q}^{\mathbf{T}}. \tag{5.9}$$

It is easy to show that this means a strategy $\mathbf{p}$ is an *Evolutionarily Stable Strategy (ESS)* for a matrix game, if and only if for any mixed strategy $\mathbf{q} \neq \mathbf{p}$

$$E[\mathbf{p}, \mathbf{p}] \geq E[\mathbf{q}, \mathbf{p}]; \tag{5.10}$$

$$\text{if } E[\mathbf{p}, \mathbf{p}] = E[\mathbf{q}, \mathbf{p}], \text{ then } E[\mathbf{p}, \mathbf{q}] > E[\mathbf{q}, \mathbf{q}], \tag{5.11}$$

(see e.g. Broom and Rychtář 2013).

We note that inequality (5.10) is the Nash equilibrium condition, but that, while necessary, it is not sufficient for stability. If (5.11) does not hold, then $\mathbf{p}$ may be invaded by a mutant that does equally well against the majority of individuals in the population (that play $\mathbf{p}$) but gets a (tiny) advantage against them by outperforming them in the (rare) contests with other mutants (playing $\mathbf{q}$).

Alternatively there is the possibility that the mutant and the residents do equally well against the mutants too. In this latter case invasion can occur by "drift"; both types do equally well, so in the absence of selective advantage random chance decides whether the frequency of mutants increases or decreases.

We define $T(\mathbf{p})$ as the set of pure strategies with equal payoffs against $\mathbf{p}$, i.e.

$$T(\mathbf{p}) = \{i : E[S_i, \mathbf{p}] = E[\mathbf{p}, \mathbf{p}]\}. \tag{5.12}$$

Theorem 1 (Bishop Cannings Theorem). *If* $\mathbf{p}$ *is an ESS of the matrix game* A *and* $\mathbf{q} \neq \mathbf{p}$ *is such that* $S(\mathbf{q}) \subseteq T(\mathbf{p})$*, then* $\mathbf{q}$ *is not an ESS of matrix game* A*.*

For a proof, see Bishop and Cannings (1976).

5.1.2.3 Dynamic Versus Static Analysis

Dynamic and static analyses are mainly complementary, however the relationship between the two is not straightforward, and there is some apparent inconsistency between the theories. Comparing the static ESS analysis and replicator dynamics, we see that the information required for each type of analysis is different. To determine whether $\mathbf{p}$ is an ESS, we need the minimum of a function

$$\mathbf{q} \to \mathscr{E}[\mathbf{p}; (1-\varepsilon)\delta_{\mathbf{p}} + \varepsilon\delta_{\mathbf{q}}] - \mathscr{E}[\mathbf{q}; (1-\varepsilon)\delta_{\mathbf{p}} + \varepsilon\delta_{\mathbf{q}}] \tag{5.13}$$

to be attained for $\mathbf{q} = \mathbf{p}$ for all sufficiently small $\varepsilon > 0$.

To understand the replicator dynamics, however, we need $\mathscr{E}[S_i; \mathbf{p}^{\mathbf{T}}]$ for all i and all $\mathbf{p}$. Thus a major difference between the two methods is that the static analysis considers monomorphic populations $\delta_{\mathbf{p}}$ while the dynamic analysis considers mixed populations $\mathbf{p}^{\mathbf{T}} = \sum_i p_i \delta_i$.

The analyses can thus produce the same (or at least similar) results only if there is an identification between $\delta_{\mathbf{p}}$ and $\mathbf{p}^{\mathbf{T}}$, as in the case of matrix games, and we note that most of the comparative analysis between the methods has assumed matrix games.

Theorem 2 (Folk Theorem of Evolutionary Game Theory, Hofbauer and Sigmund 2003). *For a matrix game with payoffs given by matrix A, we have:*

1. *If* $\mathbf{p}$ *is a Nash equilibrium, and so an ESS, of a matrix game, then* $\mathbf{p}^{\mathbf{T}}$ *is a rest point of the dynamics , i.e. the population does not evolve further from the state* $\mathbf{p}^{\mathbf{T}} = \sum_i p_i \delta_i$.
2. *If* $\mathbf{p}$ *is a strict Nash equilibrium, then* $\mathbf{p}$ *is locally asymptotically stable.*
3. *If the rest point* $\mathbf{p}^*$ *of the dynamics is also the limit of an interior orbit, then it is a Nash equilibrium.*
4. *If the rest point* $\mathbf{p}$ *is Lyapunov stable, then* $\mathbf{p}$ *is a Nash equilibrium.*

An ESS is an attractor of the replicator dynamics, and the population converges to the ESS for every strategy sufficiently close to it. If $\mathbf{p}$ is an internal ESS, then global convergence to $\mathbf{p}$ is assured (Zeeman 1980).

It is also true that if the replicator dynamics has a unique internal rest point $\mathbf{p}^*$, under certain conditions (satisfied for matrix games)

$$\lim_{t\to\infty} \frac{1}{T} \int_0^T p_i(t)dt = p_i^*, \tag{5.14}$$

so that the long-term average strategy is given by this rest point, even if there is considerable variation at any given time.

Thus for matrix games, identifying ESSs and Nash equilibria of a game gives a lot of important information about the dynamics. For example, if $\mathbf{p}$ is an internal ESS, then global convergence to $\mathbf{p}$ is assured.

However, there are cases when an ESS analysis does not provide such a complete picture. In particular, there are attractors of the replicator dynamics that are not ESSs. To see this, consider the matrix

$$\begin{pmatrix} 0 & 1 & -1 \\ -2 & 0 & 2 \\ 2 & -1 & 0 \end{pmatrix} \tag{5.15}$$

(see Hofbauer and Sigmund 1998). The replicator dynamics for this game has a unique internal attractor, but this attractor is not an ESS. This happens because we can find an invading mixture for **p** where the dynamics effectively forces the mixture into a combination that no longer invades. Thus if the invading group is comprised of mixed strategists it can invade, whereas if it is comprised of a mixture of pure strategists it cannot. Note that for the discrete dynamics the situation is even more complex, since then it is not guaranteed that an ESS is an attractor (Cannings 1990).

5.1.3 Two Classic Matrix Games

Two well-known examples of matrix games are the Hawk-Dove game (Maynard Smith et al. 1973) and the prisoner's dilemma (Tucker 1980). These both represent important biological/social scenarios.

5.1.3.1 The Hawk Dove Game

For the Hawk-Dove game, individuals compete against other randomly chosen individuals for a reward (e.g. a territory) of value $V > 0$. Each of the contestants has two pure strategies available, Hawk (H) and Dove (D). Hawks fight, whereas Doves merely display. Doves divide the reward, a Hawk always beats a Dove, whereas two Hawks fight, with the loser incurring a cost C. This gives the payoff matrix as

$$\begin{array}{c} \\ Hawk \\ Dove \end{array} \begin{array}{c} \begin{array}{cc} Hawk & Dove \end{array} \\ \begin{pmatrix} \frac{V-C}{2} & V \\ 0 & \frac{V}{2} \end{pmatrix} \end{array}. \tag{5.16}$$

Denoting a mixed strategy $\mathbf{p} = (p, 1-p)$ to mean to play Hawk with probability p and to play Dove otherwise, it is easy to show that pure Dove is never an ESS, pure Hawk is an ESS if $V \geq C$. For $V < C$, $\mathbf{p} = (V/C, 1 - V/C)$ is the unique ESS (see e.g. Broom and Rychtář 2013).

5.1.3.2 The Prisoner's Dilemma

In the Prisoner's dilemma, a pair of individuals can either cooperate (play C) or try to obtain an advantage by defecting and exploiting the other (play D). The payoffs are given by the payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} \textit{Cooperate} & \textit{Defect} \end{array} \\ \begin{array}{c} \textit{Cooperate} \\ \textit{Defect} \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix}. \end{array} \tag{5.17}$$

Whilst the individual numbers are not important, for the classical dilemma we need $T > R > P > S$. We also need the additional condition $2R > S + T$ which is necessary for the evolution of cooperation. In this game Defect is the unique ESS, although if both players cooperated they would do better. The game is widely used to consider the issue of (especially human) cooperation, and of how it can be established against cheating. Many variants of the above game, usually using multiple interactions of some kind, have been developed to this end (see e.g. Axelrod 1981; Nowak 2006).

5.2 Nonlinear Games

5.2.1 Overview and General Theory

In the previous section we considered matrix games, where

$$\mathcal{E}[\mathbf{p}; \mathbf{q}^T] = \mathbf{p}A\mathbf{q}^\mathbf{T}. \tag{5.18}$$

The above payoffs can alternatively be written in the form $\sum_i p_i(A\mathbf{q}^\mathbf{T})_i$ or $\sum_j (\mathbf{p}A)_j q_j$, and so payoffs are linear in both the strategy of the focal individual and the strategy of the population and, as we have seen, this has nice static and dynamic properties.

More generally, we say that $\mathcal{E}$ is *linear on the left* if it is linear in the strategy of the focal player, i.e.

$$\mathcal{E}\left[\sum_i \alpha_i \mathbf{p}_i; \Pi\right] = \sum_i \alpha_i \mathcal{E}[\mathbf{p}_i; \Pi] \tag{5.19}$$

for every population Π, every m-tuple of individual strategies $\mathbf{p}_1, \ldots, \mathbf{p}_m$ and every collection of constants $\alpha_i \geq 0$ such that $\sum_i \alpha_i = 1$ (Broom and Rychtář 2013).

We say that $\mathcal{E}$ is *linear on the right* if it is linear in the strategy of the population, i.e.

$$\mathcal{E}\left[\mathbf{p}; \sum_i \alpha_i \delta_{\mathbf{q}_i}\right] = \sum_i \alpha_i \mathcal{E}[\mathbf{p}; \delta_{\mathbf{q}_i}] \tag{5.20}$$

for every individual strategy $\mathbf{p}$, every m-tuple $\mathbf{q}_1, \ldots, \mathbf{q}_m$ and every collection of α_i's from [0, 1] such that $\sum_i \alpha_i = 1$ (Broom and Rychtář 2013).

Recall that for matrix games, the payoff to an individual is the same whether it faces opponents playing a polymorphic mixture of pure strategies or a monomorphic population. We say that a game has *polymorphic-monomorphic equivalence* if for every strategy $\mathbf{p}$, any finite collection of strategies $\{\mathbf{q}_i\}_{i=1}^m$ and any corresponding collection of m constants $\alpha_i \geq 0$ such that $\sum_i^m \alpha_i = 1$ we have

$$\mathscr{E}\left[\mathbf{p}; \sum_i \alpha_i \delta_{\mathbf{q}_i}\right] = \mathscr{E}\left[\mathbf{p}; \delta_{\sum_i \alpha_i \mathbf{q}_i}\right] \tag{5.21}$$

(Broom and Rychtář 2013). Note that the concept of polymorphic-monomorphic equivalence holds only in respect of static analyses, and there is no such concept in terms of dynamics.

The payoff is linear on the left for many evolutionary games because $\mathscr{E}[\mathbf{p}; \Pi]$ is often defined to be the average of the payoffs to players of pure strategy S_i, weighted by the selection probability p_i, for all i. It is common, however, that the payoff is nonlinear on the right, which occurs whenever the game does not involve pairwise contests against randomly selected opponents.

The payoff function can be nonlinear on the left, if a strategy is a pure strategy drawn from a continuum, but that the payoff is nonlinear as a function of this pure strategy, such as in the tree height game from Kokko (2007) that we consider in Sect. 5.2.4. Nearly all real situations feature nonlinearity of some type, and when models of real behaviours are developed, the payoffs involved are indeed generally nonlinear in some way.

Some results for linear games can be generalized and reformulated for nonlinear games. The conditions (5.10) and (5.11) can be generalized as follows:

Theorem 3. *For games with generic payoffs, if the incentive function*

$$h_{\mathbf{p},\mathbf{q},u} = \mathscr{E}[\mathbf{p}; (1-u)\delta_{\mathbf{p}} + u\delta_{\mathbf{q}}] - \mathscr{E}[\mathbf{q}; (1-u)\delta_{\mathbf{p}} + u\delta_{\mathbf{q}}] \tag{5.22}$$

is differentiable (from the right) at $u = 0$ for every $\mathbf{p}$ and $\mathbf{q}$, then $\mathbf{p}$ is an ESS if and only if for every $\mathbf{q} \neq \mathbf{p}$;

1. $\mathscr{E}[\mathbf{p}; \delta_{\mathbf{p}}] \geq \mathscr{E}[\mathbf{q}; \delta_{\mathbf{p}}]$ *and*
2. *if* $\mathscr{E}[\mathbf{p}; \delta_{\mathbf{p}}] = \mathscr{E}[\mathbf{q}; \delta_{\mathbf{p}}]$, *then* $\frac{\partial}{\partial u}\Big|_{(u=0)} h_{\mathbf{p},\mathbf{q},u} > 0$.

For a proof, see Broom and Rychtář (2013).

Theorem 4. *Let $\mathscr{E}$ be linear in the focal player strategy, i.e. (5.19) holds, and let the function $h_{\mathbf{p},\mathbf{q},u}$ be differentiable w.r.t u at $u = 0$. Let $\mathbf{p} = (p_i)$ be an ESS. Then $\mathscr{E}[\mathbf{p}; \delta_{\mathbf{p}}] = \mathscr{E}[S_i; \delta_{\mathbf{p}}]$ for any pure strategy S_i such that $i \in S(\mathbf{p}) = \{j; p_j > 0\}$.*

For a proof see Broom and Rychtář (2013). We note that it is enough to assume $h_{\mathbf{p},\mathbf{q},u}$ to be continuous.

If the payoff is not linear but strictly convex so that, for all $\mathbf{q}$ and all $\mathbf{p}$ with at least two elements in $S(\mathbf{p})$,

$$\sum_i p_i \mathscr{E}[S_i; \delta_{\mathbf{q}}] > \mathscr{E}[\mathbf{p}; \delta_{\mathbf{q}}], \tag{5.23}$$

then any ESS must be a pure strategy.

Lemma 1 below shows that the payoffs of games that are linear in the focal player strategy and satisfy polymorphic monomorphic equivalence (5.21) must be of a special form. These games are called *population games*, or *playing the field games*.

Lemma 1. *If the payoffs of the game are linear in the focal player strategy (i.e. satisfy (5.19)) and satisfy polymorphic monomorphic equivalence (5.21), then for every* $\mathbf{x}, \mathbf{y}, \mathbf{z}$ *and every* $\varepsilon \in [0, 1]$

$$\mathscr{E}[\mathbf{x}; (1-\varepsilon)\delta_{\mathbf{y}} + \varepsilon\delta_{\mathbf{z}}] = \sum_i x_i f_i\big((1-\varepsilon)\mathbf{y} + \varepsilon\mathbf{z}\big) \tag{5.24}$$

where $f_i(\mathbf{q}) = \mathscr{E}[S_i; \delta_{\mathbf{q}}]$.

Below we write payoffs in the form $\mathscr{E}[\mathbf{p}; \delta_{\mathbf{q}}] = \sum_i p_i f_i(\mathbf{q})$ for some functions f_i, and this indicates that payoffs are linear in the focal player strategy and also satisfy polymorphic monomorphic equivalence.

Theorem 5. *Let the payoffs be such that* $\mathscr{E}[\mathbf{p}; \delta_{\mathbf{q}}] = \sum_i p_i f_i(\mathbf{q})$ *for some continuous functions* f_i*. Then the strategy* $\mathbf{p}$ *is an ESS if and only if it is locally superior, i.e. there is* $U(\mathbf{p})$ *a neighbourhood of* $\mathbf{p}$ *such that*

$$\mathscr{E}[\mathbf{p}; \delta_{\mathbf{q}}] > \mathscr{E}[\mathbf{q}; \delta_{\mathbf{q}}], \textit{ for all } \mathbf{q}(\neq \mathbf{p}) \in U(\mathbf{p}). \tag{5.25}$$

For a proof, see Palm (1984).

5.2.2 Playing the Field

In this section we consider payoff functions of the form

$$\mathscr{E}[\mathbf{p}; \Pi] = \sum p_i f_i(\Pi) \tag{5.26}$$

where the f_i's are (in general nonlinear) functions of the population strategy Π. Such playing the field games are the most natural way of incorporating nonlinearity into a game model, since the fitness function automatically includes the population frequencies of the different strategies.

An example is the sex ratio game, one of the classical models of evolutionary game theory (Hamilton 1967). The model considers the question of why the sex ratio in most animals is close to a half? At first sight there needs to be far less males than females, since often the only male contribution is in mating; in many species most offspring are fathered by a small number of males and the rest make no contribution.

Assume that the strategy of an individual female is its choice of the proportion of male offspring. Let p be the strategy of a small invading group in a population that plays strategy m. Every individual has the same number of offspring, so fitness is given proportional to the number of grandchildren. Given that every individual has one mother and one father, if generation sizes remain constant it is easy to show that the fitness of an individual with strategy p is given by

$$\mathscr{E}[p;\delta_m] = \frac{p}{m} + \frac{1-p}{1-m} \tag{5.27}$$

so that in the notation of Eq. (5.26) we have

$$f_1(m) = \frac{1}{m},\, f_2(m) = \frac{1}{1-m}. \tag{5.28}$$

The unique ESS of this game is $m = 1/2$, i.e. an equal sex ratio. The sex ratio game is in fact effectively just a special case of the following foraging problem (with $N = 2$ and $r_1 = r_2$).

Consider a population of animals foraging on N food patches, with resources $r_i > 0$ per unit time for $i = 1, \ldots, N$, equally shared by all individuals on the patch (Parker 1978).

The game has N pure strategies for this game, each corresponding to foraging on a given patch, and a mixed strategy $\mathbf{x} = (x_i)$ means to forage at patch i with probability x_i. The payoff to an individual using strategy $\mathbf{x} = (x_i)$ against a population playing $\mathbf{y} = (y_i)$ is

$$\mathscr{E}[\mathbf{x};\delta_{\mathbf{y}}] = \begin{cases} \infty, & \text{if } x_i > 0 \text{ for some } i \text{ such that } y_i = 0, \\ \displaystyle\sum_{i;x_i>0}^{N} x_i \frac{r_i}{y_i} & \text{otherwise.} \end{cases} \tag{5.29}$$

It is clear from (5.29) that any ESS $\mathbf{p}$ must have $p_i > 0$ for all $i = 1, \ldots, N$. Thus any potential problems with infinite payoffs do not need to be considered. In particular Theorem 3 holds despite the discontinuities in the fitness functions, since they are continuous in the vicinity of any potential ESS.

The unique ESS $\mathbf{p} = (p_i)$ is given by $p_i = r_i / \sum_{i=1}^{N} r_i$. This solution can alternatively be written as

$$\frac{p_i}{p_j} = \frac{r_i}{r_j}. \tag{5.30}$$

This is called *Parker's matching principle*.

We can show this as follows. It is clear that $\mathscr{E}[\mathbf{q};\delta_{\mathbf{p}}] = \mathscr{E}[\mathbf{p};\delta_{\mathbf{p}}]$ for all $\mathbf{q}$. Moreover, since this game satisfies polymorphic monomorphic equivalence (5.21) then

$$\mathscr{E}[\mathbf{x};(1-u)\delta_{\mathbf{y}}+u\delta_{\mathbf{z}}]=\mathscr{E}[\mathbf{x};\delta_{(1-u)\mathbf{y}+u\mathbf{z}}] \tag{5.31}$$

and so

$$h_{\mathbf{p},\mathbf{q},u}=\mathscr{E}[\mathbf{p};(1-u)\delta_{\mathbf{p}}+u\delta_{\mathbf{q}}]-\mathscr{E}[\mathbf{q};(1-u)\delta_{\mathbf{p}}+u\delta_{\mathbf{q}}]= \tag{5.32}$$

$$\sum_{i=1}^{N}(p_i-q_i)\frac{r_i}{p_i+u(q_i-p_i)}= \tag{5.33}$$

$$\sum_{i=1}^{N}\frac{p_i-q_i}{p_i}r_i\left(1-u\frac{q_i-p_i}{p_i}+\dots\right). \tag{5.34}$$

This implies that

$$\frac{\partial}{\partial u}\bigg|_{u=0}h_{\mathbf{p},\mathbf{q},u}=\sum_{i=1}^{N}r_i\left(\frac{p_i-q_i}{p_i}\right)^2>0. \tag{5.35}$$

So from Theorem 3, $\mathbf{p}$ is an ESS.

5.2.3 Nonlinearity due to Non-constant Interaction Rates

Another way for nonlinear games to occur is where the strategies employed by the players affect the frequency of their interactions. The pairwise interactions may be simple, but if the strategy affects the interaction rate, then the overall payoff function can be complicated.

The simplest non-trivial scenario is a two player contest with two pure strategies S_1 and S_2, with payoffs given by the usual payoff matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{5.36}$$

but where the three types of interaction happen with probabilities not proportional to their frequencies.

Assume that each pair of S_1 individuals meet at rate r_{11}, each pair of S_1 and S_2 individuals meet at rate r_{12} and each pair of S_2 individuals meet at rate r_{22} (see Taylor and Nowak 2006). This yields the following nonlinear payoff function

$$\mathscr{E}[S_1; \mathbf{p}^\mathbf{T}] = \tfrac{ar_{11}p_1 + br_{12}p_2}{r_{11}p_1 + r_{12}p_2}, \tag{5.37}$$

$$\mathscr{E}[S_2; \mathbf{p}^\mathbf{T}] = \tfrac{cr_{12}p_1 + dr_{22}p_2}{r_{12}p_1 + r_{22}p_2}. \tag{5.38}$$

This reduces to the standard payoffs for a matrix game for the case $r_{11} = r_{12} = r_{22}$.

In the standard game with uniform interaction rates, if $a < c$ and $b > d$ there is a mixed ESS, and this is also true for non-uniform interaction rates, although the ESS proportions change. If $a > c$ and $b < d$ then there are two ESSs in the uniform case, and this is also true for non-uniform interactions, although we note that the location of the unstable equilibrium between the pure strategies changes.

Otherwise for the uniform case there is a unique ESS. For non-uniform interaction rates, there is always a single pure ESS, but sometimes there is a mixed ESS too. For $c > a > d > b$, and setting $r_{12} = 1$, this occurs if

$$r_{11}r_{22} > \left(\frac{\sqrt{(a-b)(c-d)} + \sqrt{(a-c)(b-d)}}{d-a} \right)^2. \tag{5.39}$$

The Prisoner's Dilemma is an example where $c > a > d > b$. Setting $r_{11} = r_{22} = r$ and letting $r \to \infty$ the proportion of cooperators in the mixture tends to 1 and the basin of attraction of the proportion of cooperators p in the replicator dynamics increases, tending to $p \in (0, 1]$. Thus in extreme cases, the eventual outcome of the game can be effectively the opposite to that implied by the game with uniform interaction rates.

5.2.4 Nonlinearity in the Strategy of the Focal Player

Here we consider a third case, involving games where the strategy of an individual is described by a single number (or a vector) that does not represent the probability of playing a given pure strategy, but rather represents a unique behaviour such as the intensity of a signal. We note that this is also the scenario generally considered in Adaptive Dynamics (see e.g. Metz et al. 1992; Metz 2008), though in practice stronger assumptions are generally made than we use here.

Consider the following game-theoretical model of tree growth (Koch et al. 2004; Kokko 2007). We assume that a tree has to grow large enough in order to get sunlight and not get overshadowed by neighbours; yet the more the tree grows the more of its energy has to be devoted to "standing" rather than photosynthesis.

Let $h \in [0, 1]$ be the normalized height of the tree, so that 1 is the maximum possible height of a tree. In Kokko (2007), the fitness of a tree of height h in a forest where all other trees are of height H was given by

$$\mathscr{E}[h; \delta_H] = (1 - h^3) \cdot \big(1 + \exp(H - h)\big)^{-1}, \tag{5.40}$$

where $f(h) = 1 - h^3$ represents the proportion of leaf tissue of a tree of height h and $g(h-H) = \left(1 + \exp(H-h)\right)^{-1}$ represents the advantage or disadvantage of being taller/ shorter than neighbouring trees.

What are the ESSs for the tree, i.e. the evolutionarily stable heights? Differentiating (5.40) with respect to h obtains the unique maximum for h, i.e. the best response to a given H in the population. Any ESS must be a best response to itself, and so setting $h = H$ after the above differentiation yields

$$\frac{1}{4}\left(-6H^2 + (1 - H^3)\right) = 0. \tag{5.41}$$

Equation (5.41) has only one root in (0,1) and the crossing of the x axis happens with negative derivative, so that the root is the unique ESS.

5.3 Multi-Player Games

In the previous sections we have considered games with two individuals only, or games played against "the population". We shall now consider situations with contests involving groups of individuals which are of size three or larger, selected randomly from a large population. We shall only consider multi-player matrix games (Broom et al. 1997) here. Note that another important example of a multi-player game is the multi-player war of attrition (Haigh and Cannings 1989). For an extensive review of multiplayer evolutionary games, see Gokhale and Traulsen (2014).

5.3.1 Introduction to Multi-Player Matrix Games

Consider an infinite population, from which groups of m players are selected at random to play a game. The expected payoff to an individual is obtained by simply averaging over the rewards for all possible cases, weighted by their probabilities, as for matrix games.

In general where the ordering of individuals matter, extending the bimatrix game case to m players, the payoff to each individual in position k is governed by an m-dimensional payoff matrix. However, as in matrix games, as opposed to bimatrix games, we assume that there is no significance to the ordering of the players. Thus an individual's payoff depends only upon its strategy and the combination of its opponents' strategies. We will call such games *symmetric*, and we have the following symmetry conditions:

$$a_{i_1 \ldots i_m} = a_{i_1 \sigma(i_2) \ldots \sigma(i_m)} \tag{5.42}$$

for any permutation σ of the indices $i_2, \ldots, i_m$. For the three player case, these are simply

$$a_{pqr} = a_{prq}, \text{ for all } p, q, r = 1, 2, \ldots, n. \tag{5.43}$$

The payoff to an individual playing $\mathbf{p}$ in a contest with individuals playing $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{m-1}$ respectively is written as $E[\mathbf{p}; \mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{m-1}]$. As the ordering is irrelevant, for convenience when some strategies are identical we use a power notation, for example $E[\mathbf{p}; \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3{}^{m-3}]$.

The payoffs function is given as follows

$$E[\mathbf{p}; \mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{m-1}] = \sum_{i=1}^{n} p_i \sum_{i_1=1}^{n} \cdots \sum_{i_{m-1}=1}^{n} a_{ii_1 i_2 \ldots i_{(m-1)}} \prod_{j=1}^{k-1} p_{j,i_j}, \tag{5.44}$$

where $\mathbf{p}_j = (p_{j,1}, p_{j,2}, \ldots, p_{j,n})$.

We note that, as pointed out by Gokhale and Traulsen (2010), as long as groups are selected from the population completely at random, as is usually assumed, then symmetric and non-symmetric games will have identical payoff functions. For example in the case of 3-player games, every individual is equally likely to occupy any of the ordered positions. In particular the term a_{ijk} has identical weighting to a_{ikj} in the payoff to an i-player, so that the sum of these two can be replaced by twice their average.

A multi-player matrix is super-symmetric if

$$a_{i_1 \ldots i_m} = a_{\sigma(i_1) \ldots \sigma(i_m)} \tag{5.45}$$

for any permutation σ of the indices $i_1, \ldots, i_m$.

For example, for super-symmetric three-player three strategy games, there are ten distinct payoffs. Without loss of generality we can define the three payoffs $a_{111} = a_{222} = a_{333} = 0$, and this leaves seven distinct payoffs to consider $a_{112}, a_{113}, a_{221}, a_{223}, a_{331}, a_{332}$ and a_{123}. Broom et al. (1997) considers the replicator dynamics for such games in detail, including every case where the last seven payoffs above take values of either 1 or -1. We will only discuss the simpler two strategy games here.

5.3.2 ESSs in Multi-Player Matrix Games

A strategy $\mathbf{p}$ in an m-player game is called *evolutionarily stable against a strategy* $\mathbf{q}$ if there is an $\varepsilon_{\mathbf{q}} \in (0, 1]$ such that for all $\varepsilon \in (0, \varepsilon_{\mathbf{q}}]$

$$\mathscr{E}[\mathbf{p}; (1-\varepsilon)\delta_{\mathbf{p}} + \varepsilon\delta_{\mathbf{q}}] > \mathscr{E}[\mathbf{q}; (1-\varepsilon)\delta_{\mathbf{p}} + \varepsilon\delta_{\mathbf{q}}], \tag{5.46}$$

where

$$\mathscr{E}[\mathbf{x}; (1-\varepsilon)\delta_{\mathbf{y}} + \varepsilon\delta_{\mathbf{z}}] = \sum_{l=0}^{m-1} \binom{m-1}{l} (1-\varepsilon)^l \varepsilon^{m-1-l} E[\mathbf{x}; \mathbf{y}^l, \mathbf{z}^{m-1-l}]. \tag{5.47}$$

$\mathbf{p}$ is an *ESS* for the game if for every $\mathbf{q} \neq \mathbf{p}$, there is $\varepsilon_{\mathbf{q}} > 0$ such that (5.46) is satisfied for all $\varepsilon \in (0, \varepsilon_{\mathbf{q}}]$ (Broom et al. 1997).

Similarly as in inequalities (5.10) and (5.11), we have the following:

Theorem 6. *For an m-player matrix game, the mixed strategy* $\mathbf{p}$ *is evolutionarily stable against* $\mathbf{q}$ *if and only if there is a* $j \in \{0, 1, \ldots, m-1\}$ *such that*

$$E[\mathbf{p}; \mathbf{p}^{m-1-j}, \mathbf{q}^j] > E[\mathbf{q}; \mathbf{p}^{m-1-j}, \mathbf{q}^j], \tag{5.48}$$

$$E[\mathbf{p}; \mathbf{p}^{m-1-i}, \mathbf{q}^i] = E[\mathbf{q}; \mathbf{p}^{m-1-j}, \mathbf{q}^i] \text{ for all } i < j. \tag{5.49}$$

For a proof see Broom et al. (1997) or Bukowski and Miękisz (2004).

A strategy $\mathbf{p}$ is an *ESS at level J* if, for every $\mathbf{q} \neq \mathbf{p}$, the conditions (5.48)–(5.49) of Theorem 6 are satisfied for some $j \leq J$ and there is at least one $\mathbf{q} \neq \mathbf{p}$ for which the conditions are met for $j = J$ precisely.

If $\mathbf{p}$ is an ESS, then by Theorem 6, for all q,

$$E[\mathbf{p}; \mathbf{p}^{m-1}] \geq E[\mathbf{q}; \mathbf{p}^{m-1}]. \tag{5.50}$$

The payoffs are linear on the left so that

$$E[\mathbf{p}; \mathbf{p}^{m-1}] = E[\mathbf{q}; \mathbf{p}^{m-1}], \text{ for all } \mathbf{q} \text{ with } S(\mathbf{q}) \subseteq S(\mathbf{p}). \tag{5.51}$$

We note that in the generic case, any pure ESS is of level 0. A mixed ESS cannot be of level 0, but in the generic case, any mixed ESS must be of level 1.

Analogues of the strong restrictions on possible combinations of ESSs for matrix games do not hold for multi-player games. In particular, the Bishop-Cannings Theorem fails already for $m = 3$. For $m > 3$, there can be more than one ESS with the same support as we shall see in Sect. 5.3.3. On the other hand, we still have the following for $m = 3$.

Theorem 7. *It is not possible to have two ESSs with the same support in a three player matrix game.*

For a proof, see Broom et al. (1997) or Broom and Rychtář (2013).

5.3.3 Two-Strategy Multi-Player Games

We shall now consider games with only two pure strategies. The possible situations for a given individual are thus all combinations of that individual playing pure strategy $i = 1, 2$ against $m-1$ players, j of which play strategy S_1 (and the other $m-1-j$ play strategy S_2), for any $0 \leq j \leq m-1$. We shall denote these payoffs by α_{ij}.

We consider an individual playing strategy $\mathbf{x}$ in a population playing $\mathbf{y}$. A group of $m-1$ opponents is chosen and each one of them chooses to play strategy S_1 with probability y_1 (and so strategy S_2 with probability $y_2 = 1 - y_1$). We obtain

$$\mathscr{E}[\mathbf{x}; \delta_{\mathbf{y}}] = \sum_{l=0}^{m-1} \binom{m-1}{l} y_1^l y_2^{m-1-l} E[\mathbf{x}; S_1^l S_2^{m-1-l}], \tag{5.52}$$

where

$$E[\mathbf{x}; S_1^l, S_2^{m-1-l}] = \sum_{i=1}^{2} x_i \alpha_{il}. \tag{5.53}$$

Note that it does not matter whether the population is polymorphic or monomorphic and playing the mean strategy; thus multi-player matrix games have the polymorphic-monomorphic equivalence property.

Recalling that the payoffs of the m-player two strategy matrix game are α_{il} for $i = 1, 2$ and $l = 0, 1, \ldots, m-1$, we define $\beta_l = \alpha_{1l} - \alpha_{2l}$ and consider the incentive function

$$h(p) = \mathscr{E}[S_1; \delta_{(p,1-p)}] - \mathscr{E}[S_2; \delta_{(p,1-p)}] \tag{5.54}$$

$$= \textstyle\sum_{l=0}^{m-1} \binom{m-1}{l} \beta_l p^l (1-p)^{m-l-1}. \tag{5.55}$$

The function h quantifies the benefits of using strategy S_1 over strategy S_2 in a population where all other players use strategy $\mathbf{p} = (p, 1-p)$. Note that h is differentiable, and that the replicator dynamics now becomes

$$\frac{dq}{dt} = q(1-q)h(q). \tag{5.56}$$

Theorem 8. *In a generic two strategy m-player matrix game*

1. *pure strategy S_1 is an ESS (level 0) if and only if $\beta_{m-1} > 0$,*
2. *pure strategy S_2 is an ESS (level 0) if and only if $\beta_0 < 0$,*
3. *an internal strategy $\mathbf{p} = (p, 1-p)$ is an ESS, if and only if*

1. *$h(p) = 0$, and*
2. *$h'(p) < 0$.*

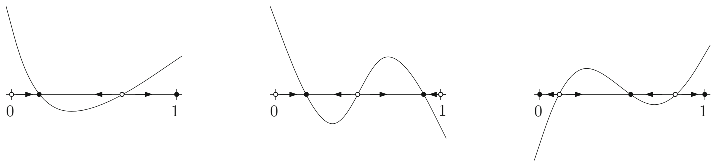

Fig. 5.2 The incentive function and ESSs in multiplayer games. The *full dots* show equilibrium points and the *arrows* show the direction of evolution under the replicator dynamics

It is shown in Broom et al. (1997) that the possible sets of ESSs are the following:

1. 0 pure ESSs, and l internal ESSs with $l \leq \lfloor \frac{m}{2} \rfloor$;
2. 1 pure ESS, and l internal ESSs with $l \leq \lfloor \frac{m}{2} - 1 \rfloor$;
3. 2 pure ESSs, and l internal ESSs with $l \leq \lfloor \frac{m}{2} - 2 \rfloor$ (Fig. 5.2).

There can be more than one ESS with the same support in a 4-player game as shown in the example below.

Consider an example with the following payoffs (Bukowski and Miękisz 2004): with $\alpha_{11} = \alpha_{22} = -\frac{13}{96}, \alpha_{13} = \alpha_{20} = -\frac{3}{32}$ and $\alpha_{10} = \alpha_{12} = \alpha_{21} = \alpha_{23} = 0$. Thus $\beta_0 = 3/32, \beta_1 = -13/96, \beta_2 = 13/96, \beta_3 = -3/32$ giving

$$h(p) = -\frac{3}{32}p^3 + \frac{13}{32}p^2(1-p) - \frac{13}{32}p(1-p)^2 + \frac{3}{32}(1-p)^3 = \tag{5.57}$$

$$-\left(p - \frac{1}{4}\right)\left(p - \frac{1}{2}\right)\left(p - \frac{3}{4}\right). \tag{5.58}$$

Using the ESS conditions from Theorem 8, we see that the game has two internal ESSs at $\mathbf{p} = (1/4, 3/4)$ and $\mathbf{p} = (3/4, 1/4)$, and no pure ESSs.

5.4 Discussion

In this paper we have considered two main recent developments in the theory of evolutionary games. In particular the extension from linear matrix games to nonlinear games, and from two player to multiplayer games.

Nonlinearity within evolutionary games is introduced in its most natural way by considering games played against the population as a whole, so-called playing the field games. These can be generally expressed in the form of Eq. (5.26). They often result from situations where individuals do not interact directly, but where their behaviours have a direct effect on the environment, which then affects the payoffs of individuals. Thus in foraging models, the value of food patches depends directly on the intensity of their use by foragers within the population, as we saw from Parker (1978). More recent and realistic models of this phenomenon are given in Cressman et al. (2004), Křivan et al. (2008) for example.

Even when games are pairwise, linearity only occurs because opponents are chosen at random, with equal probability. If some opponents are more likely than others and this is in any way related to the strategy of those involved, either through individuals directly being more likely to interact with those choosing a particular strategy or because evolution has led to different strategy distributions in different geographical locations, then nonlinearity will result, as we saw in Taylor and Nowak (2006). An example of this phenomenon occurs in food-stealing games, see e.g. Broom et al. (2004, 2008).

The above games are linear in the strategy of the focal player, as its strategy is a probabilistic weighting of distinct choices. When its strategy is a single trait chosen from a continuum, such as the height of a tree as in Koch et al. (2004), Kokko (2007), then there is nonlinearity in the focal player strategy too. Another example is the sperm allocation games of Parker et al. (1997), Ball and Parker (2007). We also note that this idea is central to the related concept of adaptive dynamics, where populations evolve by successive small mutations, see Kisdi and Meszéna (1993), Geritz et al. (1998).

Multiplayer games have been, and continue to be, common in Economics, for instance see Kim (1996),Wooders et al. (2006), Ganzfried and Sandholm (2009). However until recently they have been less common in evolutionary games. An extension of the classical idea of well-mixed populations of pairwise games to consider such populations with multiplayer games was first introduced with the work of Palm (1984) and followed by Haigh and Cannings (1989), Broom et al. (1997), Bukowski and Miękisz (2004). More recently Hauert et al. (2006), Gokhale and Traulsen (2010), Han et al. (2012), Gokhale and Traulsen (2014) have developed the theory further.

As for nonlinear games above, multiplayer games can occur from non-independent pairwise games, for example within the formation of dominance hierarchies, where the results of a contest directly dictate who an individual will face next (if anybody). This was the focus of the games from Broom et al. (2000a,b).

Evolutionary game theory has also been extended to finite populations, based upon the original Moran process (Moran 1962) where different concepts are needed to deal with the stochastics effects which are not present in infinite populations, and where the single most important concept is that of the fixation probability of a rare mutant (equivalent to a small fraction of mutants within an infinite population, whose establishment within a population is either certain or impossible), important examples include Fogel et al. (1998), Nowak et al. (2004), Taylor et al. (2004), Traulsen et al. (2005), Nowak (2006). Within this general theory, there have also been developments based upon multiplayer games, and these are well-reviewed in Gokhale and Traulsen (2014).

Interesting new work on multiplayer games in each of the above areas continues to appear. For example the theory of adaptive dynamics is continually expanding, and the nonlinearity that appeared in the food stealing games of Broom et al. (2008), which was due to the effect of time constraints, is being considered more widely, for instance in Cressman et al. (2014). The work on finite populations including its multiplayer variants continues to be developed. In particular the modelling of structured populations from evolutionary graph theory Lieberman et al. (2005) has been extended to incorporate multiplayer games (Broom and Rychtář 2012). This area is at the relatively early stages of development, and there are many possibilities for further research.

References

Axelrod R (1981) The emergences of cooperation among egoists. Am Polit Sci Rev 75:306–318
Ball M, Parker G (2007) Sperm competition games: the risk model can generate higher sperm allocation to virgin females. J Evol Biol 20(2):767–779
Bishop D, Cannings C (1976) Models of animal conflict. Adv Appl Probab 8:616–621
Bishop D, Cannings C (1978) A generalized war of attrition. J Theor Biol 70:85–124
Broom M, Rychtář J (2012) A general framework for analyzing multiplayer games in networks using territorial interactions as a case study. J Theor Biol 302:70–80
Broom M, Rychtář J (2013) Game-theoretical models in biology. CRC Press, Boca Raton, FL
Broom M, Cannings C, Vickers G (1997) Multi-player matrix games. Bull Math Biol 59(5):931–952
Broom M, Cannings C, Vickers G (2000a) Evolution in knockout conflicts: the fixed strategy case. Bull Math Biol 62(3):451–466
Broom M, Cannings M, Vickers G (2000b) Evolution in knockout contests: the variable strategy case. Selection 1(1):5–22
Broom M, Luther R, Ruxton G (2004) Resistance is useless? - extensions to the game theory of kleptoparasitism. Bull Math Biol 66(6):1645–1658
Broom M, Luther R, Ruxton G, Rychtář J (2008) A game-theoretic model of kleptoparasitic behavior in polymorphic populations. J Theor Biol 255(1):81–91
Bukowski M, Miękisz J (2004) Evolutionary and asymptotic stability in symmetric multi-player games. Int J Game Theory 33(1):41–54
Cannings C (1990) Topics in the theory of ESS's. In: Lessard S (ed) Mathematical and statistical developments of evolutionary theory. Lecture notes in mathematics, pp 95–119. Kluwer Academic Publishers, Drodrecht
Cressman R, Křivan V, Garay J (2004) Ideal free distributions, evolutionary games, and population dynamics in multiple-species environments. Am Nat 164(4):473–489
Cressman R, Křivan V, Brown J, Garay G (2014) Game-theoretic methods for functional response and optimal foraging behavior. PLoS One 9(2):e88773. doi:10.1371/journal.pone.0088773
Edwards A (2000) Foundations of mathematical genetics. Cambridge University Press, Cambridge
Fisher R (1930) The genetical theory of natural selection. Clarendon Press, Oxford
Fogel GB, Andrews PC, Fogel DB (1998) On the instability of evolutionary stable strategies in small populations. Ecol Model 109(3):283–294
Ganzfried S, Sandholm TW (2009) Computing equilibria in multiplayer stochastic games of imperfect information. In: Proceedings of the 21st international joint conference on artificial intelligence (IJCAI)
Geritz S, Kisdi E, Meszéna G, Metz J (1998) Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree. Evol Ecol 12:35–57
Gokhale C, Traulsen A (2010) Evolutionary games in the multiverse. Proc Natl Acad Sci 107(12):5500–5504
Gokhale CS, Traulsen A (2014) Evolutionary multiplayer games. Dyn Games Appl 4(4):468–488
Haigh J, Cannings C (1989) The n-person war of attrition. Acta Applicandae Math 14(1):59–74
Hamilton W (1967) Extraordinary sex ratios. Science 156:477–488
Han TA, Traulsen A, Gokhale CS (2012) On equilibrium properties of evolutionary multi-player games with random payoff matrices. Theor Popul Biol 81(4):264–272

Hauert C, Michor F, Nowak MA, Doebeli M (2006) Synergy and discounting of cooperation in social dilemmas. J Theor Biol 239(2):195–202

Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University Press, Cambridge

Hofbauer J, Sigmund K (2003) Evolutionarily game dynamics. Bull Am Math Soc 40(4):479–519

Kim Y (1996) Equilibrium selection in *n*-person coordination games. Games Econ Behav 15(2):203–227

Kisdi É, Meszéna G (1993) Density dependent life history evolution in fluctuating environments. Lect Notes Biomath 98:26–62

Koch G, Sillett S, Jennings G, Davis S (2004) The limits to tree height. Nature 428(6985):851–854

Kokko H (2007) Modelling for field biologists and other interesting people. Cambridge University Press, Cambridge

Křivan V, Cressman R, Schneider C (2008) The ideal free distribution: a review and synthesis of the game-theoretic perspective. Theor Popul Biol 73(3):403–425

Lieberman E, Hauert C, Nowak M (2005) Evolutionary dynamics on graphs. Nature 433(7023):312–316

Maynard Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge

Maynard Smith J, Price G (1973) The logic of animal conflict. Nature 246:15–18

Metz J (2008) In: SE Jørgensen, Fath BD (eds) Fitness. Encyclopedia of ecology, Five-Volume Set, 1st Edition published by Elsevier, pp 1599–1612. http://store.elsevier.com/product.jsp?isbn=9780080454054

Metz J, Nisbet R, Geritz S (1992) How should we define fitness for general ecological scenarios? Trends Ecol Evol 7(6):198–202

Moran P (1962) The statistical processes of evolutionary theory. Clarendon Press/Oxford University Press, Oxford

Nowak M (2006) Evolutionary dynamics, exploring the equations of life First Edition, Belknap Press 2006, ISBN-13: 978-0674023383

Nowak MA, Sasaki A, Taylor C, Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. Nature 428(6983):646–650

Palm G (1984) Evolutionary stable strategies and game dynamics for n-person games. J Math Biol 19(3):329–334

Parker G (1978) Searching for mates. In: Krebs JR, Davies NB (eds) Behavioural ecology: an evolutionary approach. Blackwell, Oxford, pp 214–244

Parker G, Ball M, Stockley P, Gage M (1997) Sperm competition games: a prospective analysis of risk assessment. Proc R Soc Lond Ser B Biol Sci 264(1389):1793–1802

Samuelson L (1997) Evolutionary games and equilibrium selection. MIT Press, Cambridge

Taylor P, Jonker L (1978) Evolutionarily stable strategies and game dynamics. Math Biosci 40:145–156

Taylor C, Nowak M (2006) Evolutionary game dynamics with non-uniform interaction rates. Theor Popul Biol 69(3):243–252

Taylor C, Fudenberg D, Sasaki A, Nowak M (2004) Evolutionary game dynamics in finite populations. Bull Math Biol 66(6):1621–1644

Traulsen A, Claussen JC, Hauert C (2005) Coevolutionary dynamics: from finite to infinite populations. Phys Rev Lett 95(23):238,701

Tucker A (1980) On Jargon: the Prisoner's dilemma. UMAP J 1:101

Wooders M, Cartwright E, Selten R (2006) Behavioral conformity in games with many players. Games Econ Behav 57(2):347–360

Zeeman E (1980) Population dynamics from game theory. In: Nitecki Z, Robinson C (eds) Global theory of dynamical systems. Lecture notes in mathematics, vol 819. Springer, Berlin

Chapter 6
A Zero-Sum Game Between the Network Designer and an Adversary in Consensus Protocols

Mahmoud El Chamie and Tamer Başar

Abstract This article addresses the problem of designing optimal strategies in consensus protocols for networks vulnerable to adversarial attacks. First, a set of necessary conditions for optimal control is given in the case of the dynamic (multi-stage) weight selection problem of consensus protocols. Under some mild conditions, it turns out that only one-stage is sufficient for reaching consensus, and the article derives a closed-form solution for the optimal control. Second, a (zero-sum) game theoretical model with a "convex-convex" quadratic objective function is considered for the problem of a network with an adversary corrupting the control signal with noise. Mixed-strategy saddle-point (MSSP) strategies are obtained for the players (the adversary and the network designer) in the resulting game. Further, a totally distributed gradient method that computes the optimal control is provided. Simulation results show that an adversary using an MSSP strategy can drive the system away from consensus, while an adversary using a uniform random strategy does not cause as much damage.

Keywords Consensus protocols • Adversary • Convex-convex zero-sum quadratic games • Saddle-point strategies • Distributed control

---

M. El Chamie (✉)
INRIA Sophia Antipolis-Méditerranée, 2004 route des Lucioles - BP 93, 06902 Sophia Antipolis Cedex, France
e-mail: shamieh.m@gmail.com

T. Başar
Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 West Main, Urbana, IL 61801, USA
e-mail: basar1@illinois.edu

© Springer International Publishing Switzerland 2016

F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_6

6.1 Introduction

Consensus protocols are attracting increasing interest in recent years. Consensus algorithms are based on neighbor-to-neighbor interactions among nodes, where each node executes iteratively a weighted average linear update rule, and the goal is for the entire set of nodes to reach consensus. Many applications such as formation control (Fax and Murray 2004), load balancing (Cybenko 1989), distributed state estimation in power systems (Vukovic and Dan 2014), and data fusion in sensor networks (Avrachenkov et al. 2011) rely on such protocols. Consensus in networks can be subject to time-varying network topology (Ren and Beard 2005), quantization in communication (Nedic et al. 2009; El Chamie et al. 2014), communication delays (Olfati-Saber and Murray 2004), and adversarial intervention (Khanafer et al. 2013).

As in any protocol, some parameters can be tuned in the consensus algorithm. Therefore optimizing the choice of these parameters leads to a better performance in terms of energy savings, speed of convergence, or robustness of the system. For instance, energy savings can be achieved by applying termination algorithms of the consensus protocol. El Chamie et al. (2013) give distributed algorithms to reduce communication overhead and (Ko and Gao 2009; Hendrickx et al. 2014) study finite-time consensus using arbitrary time-varying weights selected before the start of consensus that are based on matrix factorization techniques. For achieving a faster asymptotic convergence rate using a *fixed* weight selection algorithm, the weights in consensus protocols can be tuned in centralized manner by a semi-definite program (SDP) (Xiao and Boyd 2004) or in distributed manner using projected sub-gradient methods (El Chamie et al. 2015). For *time-varying* weights, the work in (Schwarz and Matz 2012) selects the weights to reduce the mean square error in correlated or uncorrelated initial node values. For a complete overview of consensus protocols, we refer the reader to El Chamie (2014) and Olfati-Saber et al. (2007), and the references therein.

Further, networks can be susceptible to attacks from adversaries willing to drive the system away from consensus. There are different types of adversaries depending on their action strategies. For example, compromised strategic nodes [like faulty nodes or stubborn ones (Acemoglu et al. 2011; Ben-Ameur et al. 2012)] are inside-system adversaries that have access to part of the physical network. Other types of strategic intervention include adversaries that cut communication links or insert noise signals in the agents' interaction protocol (Khanafer et al. 2013). Adversaries can also inject false data (collected by nodes) into the system, which bypass bad-data detection mechanisms. False data injections are known as stealth attacks and are widely studied in problems of security of state estimation in electric power networks (Vukovic and Dan 2014; Liu et al. 2009). In order to mitigate the effect of an adversary, security procedures should be taken into account in the design of optimal strategies in consensus protocols.

Our present work shares with this set of references the same objectives of designing time-varying weights for faster consensus and studying optimal strategies in networks that are vulnerable to adversarial attacks. We study time-varying weights for consensus protocols within the framework an optimal control formulation. We then study the effect of adversaries that can compromise these weights. We propose a game theoretical framework for an adversary that can add noise to the weights to drive the system away from consensus. We derive the optimal strategies for both players (the adversary and the network designer) within the solution concept of mixed-strategy saddle-point (MSSP) equilibrium. The difference with previous related work is that in this article we consider the initial values as input to our *dynamic* weight design. This article is an extension of our work reported in the conference paper (El Chamie and Başar 2014); besides providing an overview of the results in El Chamie and Başar (2014), we consider here a distributed implementation of the optimal control using gradient methods and we further carry out extensive simulations to corroborate the theoretical findings.

6.2 Problem Formulation

In a nutshell, a network is comprised of nodes (or agents) and links that connect these nodes within a graph-theoretic topology. In this article we consider the links to be *communication links*, which allow the nodes to share information and resources. Let there be n nodes in the network, where each one has a scalar $x_i(k) \in \mathbb{R}$ called node's *state variable* that is located (and can be updated) in its local memory, where k is a discrete-time index and $x_i(0)$ is the initial value at node i. Average consensus protocol is an iterative process where nodes, subject to some given communication constraints, reach consensus on the average of all initial values (i.e., they all end up with the value $x_{ave} := \frac{1}{n}\sum_i x_i(0)$). The communication links in the network could be uni-directional (i.e., information can flow only in one direction) or bi-directional (i.e., information is allowed to flow in both directions); here we adopt the latter. That is, we model the network as an *undirected connected* graph $G = (V, E)$ where $V = \{1, \ldots, n\}$ is the set of vertices (nodes) and $E = \{1, \ldots, m\}$ is the set of edges (links). We use the notation $s \sim (ij)$ to indicate that the vertices i and j are incident to link s. One class of algorithms to achieve consensus is obtained by nodes updating their values in a synchronized and iterative way as follows:

$$\mathbf{x}_{k+1} = W(k)\mathbf{x}_k, \tag{6.1}$$

where $\mathbf{x}_k$ is the state vector having $x_i(k)$ for $i = 1, \ldots, n$ as its elements and $W(k)$ is the weight matrix at iteration k satisfying, for its ij'th element, $w_{ij} = 0$ if $(ij) \notin E$. The values of the state variables at the nodes are guaranteed to converge asymptotically to the average (under some conditions on the weights $W(k)$),

$$\lim_{k\to\infty} \mathbf{x}_k = \bar{\mathbf{x}},$$

where $\bar{\mathbf{x}} = x_{ave}\mathbf{1}$ and $\mathbf{1}$ is the vector of all ones. One such set of conditions with fixed weights (i.e., $W(k) = W \ \forall k$) is the following (Xiao and Boyd 2004):

$$\mathbf{1}^T W = \mathbf{1}^T, \ \ W\mathbf{1} = \mathbf{1}, \ \ \rho(W - \frac{1}{n}\mathbf{1}\mathbf{1}^T) < 1,$$

where $\rho(.)$ is the largest eigenvalue in magnitude of a matrix. Let us now provide some insight about these conditions. By the first condition, the average in the network is conserved, namely

$$\mathbf{1}^T\mathbf{x}_k = \mathbf{1}^T\mathbf{x}_0 = nx_{ave} \ \forall k. \tag{6.2}$$

The second ensures stability (i.e., if the system reached consensus at a given iteration, then the values of the nodes' variables will be stable and would not change in further iterations). The last condition guarantees contraction on the weight matrix (i.e., the variables eventually converge to consensus). At any iteration k, the squared error L_k from consensus is defined as follows:

$$L_k = ||\mathbf{x}_k - \bar{\mathbf{x}}||_2^2 = \mathbf{y}_k^T\mathbf{y}_k, \tag{6.3}$$

where $\mathbf{y}_k = \mathbf{x}_k - \bar{\mathbf{x}}$.

In this article, we design time-varying weight matrices $W(k)$ such that consensus is reached with the least number of iterations (that is, we are interested in achieving fastest convergence) under the criterion of minimum squared error. In this work, as opposed to others in the literature, the weight matrix is a function of both the network structure and the initial values, i.e.,

$$W(k) = W(k, \mathbf{x}_0).$$

As the weight matrix depends on initial values, a centralized unit (such as the network designer) is assumed to have a global knowledge on the network structure and on these values. In the article, we will also discuss a decentralized design of this weight matrix by using gradient methods that converge to the desired control values.

6.3 Optimal Weight Selection on Undirected Graphs

Toward the goal stated above, as commonly assumed on undirected graphs, we impose the following properties on the weight matrix:

$$W(k) = W(k)^T \text{ and } W(k)\mathbf{1} = \mathbf{1}. \tag{6.4}$$

With these conditions, the average is conserved with every iteration (i.e., Eq. (6.2) is satisfied). Let $\mathbf{u}_k \in \mathbb{R}^m$ be the control variable. At stage k, the network designer will select a control $\mathbf{u}_k$ where each element in this vector corresponds to the weight of a link in the graph at a given iteration. By considering the equality constraints in Eq. (6.4), the weight matrix can be written as a function of the control vector as follows:

$$W(k) = I_n - Q\text{diag}(\mathbf{u}_k)Q^T, \tag{6.5}$$

where I_n is the n by n identity matrix, Q is an $n \times m$ incidence matrix of the graph G (each column corresponds to an edge such that if column $s \sim (ij) \in E$, then $Q_{is} = +1$ and $Q_{js} = -1$ while all other elements of the column are zeros).

For any iteration k, the system deviation from the average can be measured by the squared error L_k. Since the goal is to reach consensus fast, in the criterion to be optimized, the error is considered only at the last stage. The problem is then to select a control that minimizes L_N, for some pre-selected N. Let $J_N := \mathbf{x}_N^T \mathbf{x}_N$, and note that

$$\begin{aligned} L_N &= \mathbf{y}_N^T \mathbf{y}_N \\ &= \mathbf{x}_N^T \mathbf{x}_N - 2\bar{\mathbf{x}}^T \mathbf{x}_N + \bar{\mathbf{x}}^T \bar{\mathbf{x}} \\ &= J_N - 2x_{ave}\mathbf{1}^T \mathbf{x}_N + n x_{ave}^2 \\ &= J_N - n x_{ave}^2. \end{aligned}$$

Then any optimal control that minimizes L_N minimizes also the function J_N because the term nx_{ave}^2 depends only on the initial values (through the average). In fact, J_N can be viewed as a cost on the system due to the control applied. The optimal control problem can then be formulated as follows:

$$\begin{aligned} &\underset{\mathbf{u}_0,\ldots,\mathbf{u}_{N-1}}{\text{minimize}} \quad J_N \\ &\text{subject to} \\ &\mathbf{x}_{k+1} = \mathbf{x}_k - Q\text{diag}(\mathbf{u}_k)Q^T \mathbf{x}_k, \ \text{for } k = 0, \ldots, N-1, \end{aligned} \tag{6.6}$$

where N is the (fixed) number of stages in this optimization. We first show that an optimal control exists.

6.3.1 Existence of a Solution

To show that an optimal control $(\mathbf{u}_k^*,\ k = 0, \ldots, N-1)$ exists, we first re-write the optimal control problem (6.6) as an unconstrained optimization problem:

$$\underset{\mathbf{u}_0,\ldots,\mathbf{u}_{N-1}}{\text{minimize}} \ f(\mathbf{u}_0, \ldots, \mathbf{u}_{N-1}) \tag{6.7}$$

where

$$f(\mathbf{u}_0, \ldots, \mathbf{u}_{N-1}) = J_N = \mathbf{x}_N^T \mathbf{x}_N$$
$$= \mathbf{x}_0^T U_{(N-1,0)}^T U_{(N-1,0)} \mathbf{x}_0, \quad (6.8)$$

and $U_{(N-1,0)} = W(N-1)W(N-2)\ldots W(0)$. Notice that the elements of the matrix $U_{(N-1,0)}$ are linear in the control variables, and $U_{(N-1,0)}^T U_{(N-1,0)}$ is a positive semi-definite matrix. Then $f(.)$ is a quadratic function and bounded from below. Hence there exists at least one control vector sequence $(\mathbf{u}_k^*,\ k = 0, \ldots, N-1)$ that globally minimizes f. As both problems (6.6) and (6.7) are equivalent, then the existence of a solution for (6.7) as demonstrated in this part guarantees a solution for (6.6).

6.3.2 Necessary Conditions

To find necessary conditions for the optimal control, we apply the maximum principle (Lewis et al. 2012, p. 24) to problem (6.6). For $k = 0, \ldots, N-1$, the system equation, performance index, and Hamiltonian are given as:

- System equation:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - Q\text{diag}(\mathbf{u}_k)Q^T\mathbf{x}_k, \quad (6.9)$$

- Performance index:

$$J_N = \mathbf{x}_N^T \mathbf{x}_N,$$

- Hamiltonian:

$$H^k = \lambda_{k+1}^T \left(\mathbf{x}_k - Q\text{diag}(\mathbf{u}_k)Q^T\mathbf{x}_k\right), \quad (6.10)$$

  where λ_{k+1} is the costate variable corresponding to iteration k.

Then, the costate equation and the associated boundary condition are:

- Costate:

$$\lambda_k = \frac{\partial H^k}{\partial \mathbf{x}_k} = \left(I_n - Q\text{diag}(\mathbf{u}_k)Q^T\right)\lambda_{k+1}, \quad (6.11)$$

- Boundary condition: $\lambda_N = \mathbf{x}_N$.

By Pontryagin's minimum principle (also known as the maximum principle) any optimal control should minimize the Hamiltonian (Lewis et al. 2012). This provides necessary conditions for an optimum control because it should satisfy this principle along any optimum trajectory. Notice that the Hamiltonian in Eq. (6.10) can be written as follows:

$$\begin{aligned} H^k &= \lambda_{k+1}^T \left(\mathbf{x}_k - Q\mathrm{diag}(\mathbf{u}_k)Q^T\mathbf{x}_k\right), \\ &= \lambda_{k+1}^T\mathbf{x}_k - \lambda_{k+1}^T Q\mathrm{diag}(Q^T\mathbf{x}_k)\mathbf{u}_k, \end{aligned} \tag{6.12}$$

and hence H^k is linear in the *unconstrained* control variables $\mathbf{u}_k$. If any coefficient of a control variable in (6.12) is nonzero, then a control that minimizes H^k would be unbounded because it is unconstrained. But an optimal control exists as we have already shown, so by applying the maximum principle, all the coefficients of the control variables in (6.12) are necessarily equal to zero, i.e.,

$$\frac{\partial H^k}{\partial \mathbf{u}_k} = \left(Q^T\mathbf{x}_k\right) \odot \left(Q^T\lambda_{k+1}\right) = \mathbf{0}, \quad \text{for } k = 0, \dots, N-1, \tag{6.13}$$

where $\odot$ is the element-wise product of the vectors and $\mathbf{0}$ is the vector of all zeros. Equation (6.13) provides necessary conditions for a controller to minimize (6.8) [and equivalently, these conditions are necessary for any optimal controller of problem (6.6)].

When $N = 1$, the boundary condition gives $\lambda_1 = \mathbf{x}_1$, and then the necessary conditions (6.13) would reduce to, $\left(Q^T\mathbf{x}_0\right) \odot \left(Q^T\mathbf{x}_1\right) = \mathbf{0}$, i.e.,

$$(x_i(0) - x_j(0))(x_i(1) - x_j(1)) = 0 \quad \text{for all } (ij) \in E. \tag{6.14}$$

Let us provide a graphical interpretation of these conditions. Let $G' = (V, E')$ be a sub-graph of G defined on the same set of vertices, V, and with links $E' \subseteq E$ such that $(ij) \in E'$ if $(ij) \in E$ and $x_i(0) - x_j(0) \neq 0$. Then we have:

Proposition 1 *If $G' = (V, E')$ is connected, then any optimal control $\mathbf{u}^*$ drives the system to consensus in one iteration, i.e.,*

$$\bar{\mathbf{x}} = \left(I_n - Q\mathit{diag}(\mathbf{u}^*)Q^T\right)\mathbf{x}_0.$$

Proof. From (6.14), $x_i(1) = x_j(1) \quad \forall (ij) \in E'$. If G' is connected, then there is a path in E' between any two vertices, and thus $x_i(1) = x_j(1)\ \forall i, j \in V$. Using also the fact that the average is conserved [by (6.2)], we get $x_i(1) = x_{ave} \quad \forall i \in V$. □

Notice that the assumption $G' = (V, E')$ being connected is not restrictive. Without any loss of generality, we can assume that this condition is satisfied. For example, if the node's initial values $x_i(0),\ i = 1, \dots, n$, are i.i.d. continuous random variables, then G' is connected almost surely because G is connected. This condition can also be satisfied almost surely by a distributed pre-processing operation by the nodes. For example the nodes can (1) add a random value to nodes' initial values, (2) perform one averaging iteration, and (3) subtract back the random value added in step (1). This procedure will not change the average value x_{ave} but at the same time it gives new initial values satisfying the connectivity assumption almost surely. In the rest of this article, we will assume that G' is connected, and therefore only one stage ($N = 1$) is needed for the operation to converge to the average.

Remark. The controls considered in this article are unconstrained. However, due to the structure of the problem, the state is constrained. In particular, for any stage k, $\mathbf{1}^T\mathbf{x}_k = \mathbf{1}^T\mathbf{x}_0$ as Eq. (6.2) demonstrates. In the presence of additional constraints on the control, the problem would be much more challenging and the control derived in this article can no longer have closed-form solutions. However, the maximum principle can still reveal some structure for the optimal solution. Suppose, for example, that $N = 1$ and that the control variables are restricted to nonnegative values (i.e., $\mathbf{u} \geq 0$). Since the Hamiltonian in Eq. (6.10) is linear in the control, if the coefficient of a control u_l is positive, u_l^* would necessarily be equal to 0. Otherwise, Eq. (6.14) must be satisfied. Therefore, for any link $l \sim (ij)$ such that $x_i(0) - x_j(0) \neq 0$, we have either $u_l^* = 0$ or $x_i(1) = x_j(1)$. This results in clusters of nodes, where each cluster has nodes with equal state variables after one iteration. But there are no simple conditions on the initial state for the cluster to include all the nodes as Proposition 1 shows for the case of unconstrained control. □

6.4 Closed-Form Solution for the One-Stage Problem

When $N = 1$, the control is a single vector $\mathbf{u}$ where each component is the weight for the corresponding edge. The optimization problem in this case is the following:

$$\mathbf{u}_S = \underset{\mathbf{u}}{\operatorname{argmin}} \quad f(\mathbf{u}), \tag{6.15}$$

where $\mathbf{u}_S$ is the solution set (possibly an infinite set) and

$$\begin{aligned} f(\mathbf{u}) &= \mathbf{x}_0^T(I_n - Q\mathrm{diag}(\mathbf{u})Q^T)(I_n - Q\mathrm{diag}(\mathbf{u})Q^T)\mathbf{x}_0 \\ &= ||\mathbf{x}_0 - Q\mathrm{diag}(\mathbf{u})Q^T\mathbf{x}_0||_2^2 \\ &= ||\mathbf{x}_0 - Q\mathrm{diag}(Q^T\mathbf{x}_0)\mathbf{u}||_2^2 \\ &= ||D\mathbf{u} - \mathbf{x}_0||_2^2, \end{aligned}$$

where

$$D = Q\mathrm{diag}(Q^T\mathbf{x}_0). \tag{6.16}$$

The problem is then convex and is reduced to a least squares approximation problem, where any element in the solution set $\mathbf{u}_S$ satisfies what is known as the normal equations:

$$D^T D\mathbf{u} = D^T \mathbf{x}_0, \quad \forall \mathbf{u} \in \mathbf{u}_S. \tag{6.17}$$

Moreover, $\mathbf{u}_S$ is not empty, with at least one solution, given by

$$\hat{\mathbf{u}} = D^+ \mathbf{x}_0, \tag{6.18}$$

where D^+ is the pseudo inverse of D that can be obtained using the singular value decomposition of D. If $D^T D$ is a positive-definite matrix, then $D^+ = (D^T D)^{-1} D^T$ and $\hat{\mathbf{u}}$ is the unique solution to the least squares problem. Let us study in more detail the singularity property of the matrix $D^T D$. We have that

$$\begin{aligned} \text{rank}(D^T D) &= \text{rank}(D) \\ &\leq \text{rank}(Q) \\ &\leq n - 1, \end{aligned}$$

where the last equality is due to the fact that the rows are not linearly independent because the sum of all rows in Q is equal to $\mathbf{0}^T$ where $\mathbf{0}$ is the vector of all zeros, so that $\text{rank}(Q) = \text{row rank}(Q) \leq n - 1$. Since $D^T D$ is an m by m matrix, for this matrix to be non-singular it is necessary that $m \leq n - 1$. But m is the number of links in the network, so a necessary condition for the matrix to be non-singular is to have a cycle-free graph (i.e., the graph G has a tree topology) where $m = n - 1$. In fact, the solution set $\mathbf{u}_S$ can be characterized by the following expression:

$$\mathbf{u} = D^+ \mathbf{x}_0 + (I_m - D^+ D)\mathbf{e},$$

where $\mathbf{e} \in \mathbb{R}^m$ is any vector. Notice that the second term in the sum is a vector that belongs to the null space of $D^T D$, and therefore with any vector $\mathbf{e}$, the control vector $\mathbf{u}$ satisfies the normal equations in (6.17). The solution $\hat{\mathbf{u}}$ is obtained by taking $\mathbf{e} = \mathbf{0}$, and it is worth noting that $\hat{\mathbf{u}}$ has the minimum L_2-norm in $\mathbf{u}_S$, i.e., $\hat{\mathbf{u}} = \text{argmin}_{\mathbf{u} \in \mathbf{u}_S} ||\mathbf{u}||$.

We denote by S the minimum value of the function $f(\mathbf{u})$:

$$S = f(\hat{\mathbf{u}}) = ||(DD^+ - I)\mathbf{x}_0||_2^2, \tag{6.19}$$

which we will have occasion to use later.

6.5 Network with an Adversary

Networks can be susceptible to attacks from adversaries. In this section, we consider an adversary that can inject noise onto the weights of the links, with the objective to drive the system away from consensus. Considering the one stage optimization

($N = 1$), the state equation would then become

$$\begin{aligned} \mathbf{x}_1 &= W(\mathbf{u}, \mathbf{v})\mathbf{x}_0 \\ &= (I_n - Q\text{diag}(\mathbf{u} + \mathbf{v})Q^T)\mathbf{x}_0, \end{aligned} \tag{6.20}$$

where $W(\mathbf{u}, \mathbf{v})$ is the weight matrix that depends on the control $\mathbf{u} \in U_1 = \mathbb{R}^m$ and the noise of the adversary $\mathbf{v} \in U_2 = \{\mathbf{y}; \mathbf{y} \in \mathbb{R}^m, ||\mathbf{y}|| \leq C\}$, where C is a given positive constant and can be interpreted as the power constraint of the adversary (the larger C, the more powerful is the adversary). The cost function is now

$$\begin{aligned} J(\mathbf{u}, \mathbf{v}) &= \mathbf{x}_1^T \mathbf{x}_1 \\ &= ||(I_n - Q\text{diag}(\mathbf{u} + \mathbf{v})Q^T)\mathbf{x}_0||_2^2 \\ &= ||D(\mathbf{u} + \mathbf{v}) - \mathbf{x}_0||_2^2, \end{aligned} \tag{6.21}$$

where D is given by (6.16). The adversary ($\mathbf{v}$) is the maximizer of $J(\mathbf{u}, \mathbf{v})$ while the network designer ($\mathbf{u}$) is the minimizer in this zero-sum two-person game having a "convex-convex" quadratic objective function because $J(\mathbf{u}, \mathbf{v})$ is convex in $\mathbf{u}$ and is convex in $\mathbf{v}$ as well.

Definition 1. A pair ($\mathbf{u}^* \in U_1, \mathbf{v}^* \in U_2$) is a pure-strategy saddle point (PSSP) of $J(\mathbf{u}, \mathbf{v})$ if the following holds:

$$J(\mathbf{u}^*, \mathbf{v}) \leq J(\mathbf{u}^*, \mathbf{v}^*) \leq J(\mathbf{u}, \mathbf{v}^*), \text{ for all } (\mathbf{u} \in U_1, \mathbf{v} \in U_2).$$

From Definition 1, any saddle-point pair ($\mathbf{u}^* \in U_1, \mathbf{v}^* \in U_2$) in pure strategies for the zero-sum game satisfies the following property (Başar and Olsder 1999):

$$J(\mathbf{u}^*, \mathbf{v}^*) = \sup_{\mathbf{v} \in U_2} \inf_{\mathbf{u} \in U_1} J(\mathbf{u}, \mathbf{v}) = \inf_{\mathbf{u} \in U_1} \sup_{\mathbf{v} \in U_2} J(\mathbf{u}, \mathbf{v}). \tag{6.22}$$

As J is a quadratic function of $\mathbf{u}$, and $J(\mathbf{u}, \mathbf{v}) \geq 0$ for all ($\mathbf{u} \in U_1, \mathbf{v} \in U_2$), for any given $\mathbf{v} \in U_2$, J attains a minimum on U_1 (Hildebrandt 1908). Moreover, since U_2 is compact, and J is a continuous function on its domain of definition, for any given $\mathbf{u} \in U_1$, J attains a maximum on U_2 by the Weierstrass Theorem. Therefore, we can replace $\inf_{\mathbf{u} \in U_1}$ by $\min_{\mathbf{u} \in U_1}$ and $\sup_{\mathbf{v} \in U_2}$ by $\max_{\mathbf{v} \in U_2}$ in (6.22). In the sequel, we will show that actually in the formulated zero-sum game we have $\max_{\mathbf{v} \in U_2} \min_{\mathbf{u} \in U_1} J(\mathbf{u}, \mathbf{v}) < \min_{\mathbf{u} \in U_1} \max_{\mathbf{v} \in U_2} J(\mathbf{u}, \mathbf{v})$, that is strict inequality holds and hence the game does not have a saddle point (in pure strategies). It, however, admits a mixed-strategy saddle-point solution (shortly to be defined and verified).

6.5.1 The Max-Min Solution

In the max-min solution, the network designer has access to the choice of the adversary:

$$\operatorname*{argmin}_{\mathbf{u}} J(\mathbf{u}, \mathbf{v}) = \operatorname*{argmin}_{\mathbf{u}} ||D(\mathbf{u} + \mathbf{v}) - \mathbf{x}_0||_2^2 = D^+\mathbf{x}_0 - \mathbf{v}.$$

Then we have,

$$\max_{\mathbf{v}} \min_{\mathbf{u}} J(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{v}} J(D^+\mathbf{x}_0 - \mathbf{v}, \mathbf{v}) = \max_{\mathbf{v}} S = S,$$

where S is the value of the one player optimization problem, given by (6.19), and is independent of $\mathbf{v}$. In other words, if the network designer knows exactly what the strategy of the adversary is (by knowing $\mathbf{v}$), then it is possible to tune his unconstrained optimal control $\mathbf{u}$ so that it eliminates the effect of the added noise vector by the adversary.

6.5.2 The Min-Max Solution

In the min-max solution, the adversary has access to the choice of the controller. Note that J can be written as:

$$\begin{aligned} J(\mathbf{u}, \mathbf{v}) &= ||D(\mathbf{u} + \mathbf{v}) - \mathbf{x}_0||_2^2 \\ &= \mathbf{x}_0^T\mathbf{x}_0 + \mathbf{u}^T D^T D\mathbf{u} - 2\mathbf{x}_0^T D\mathbf{u} \\ &\quad + \mathbf{v}^T D^T D\mathbf{v} + 2\mathbf{v}^T \left(D^T D\mathbf{u} - D^T\mathbf{x}_0\right). \end{aligned}$$

Consider the following strategy $\mathbf{v}_1$ by the adversary:

$$\begin{cases} \mathbf{v}_1 \in \mathscr{R}(D^T D) \cap U_2 & \text{if } D^T D\mathbf{u} - D^T\mathbf{x}_0 = \mathbf{0} \\ \mathbf{v}_1 = C\frac{(D^T D\mathbf{u} - D^T\mathbf{x}_0)}{||D^T D\mathbf{u} - D^T\mathbf{x}_0||} & \text{otherwise,} \end{cases} \tag{6.23}$$

where $\mathscr{R}(D^T D)$ is the range of the matrix $D^T D$. Therefore,

$$\begin{aligned} \min_{\mathbf{u}} \max_{\mathbf{v}} J(\mathbf{u}, \mathbf{v}) &\geq \min_{\mathbf{u}} J(\mathbf{u}, \mathbf{v}_1) \\ &= \min_{\mathbf{u}} \Big\{ \underbrace{\mathbf{v}_1^T D^T D\mathbf{v}_1 + 2\mathbf{v}_1^T \left(D^T D\mathbf{u} - D^T\mathbf{x}_0\right)}_{>0 \text{ due to (6.23)}} \end{aligned}$$

$$+ \mathbf{x}_0^T\mathbf{x}_0 + \mathbf{u}^T D^T D\mathbf{u} - 2\mathbf{x}_0^T D\mathbf{u}\Big\}$$
$$> \min_{\mathbf{u}} \left\{\mathbf{x}_0^T\mathbf{x}_0 + \mathbf{u}^T D^T D\mathbf{u} - 2\mathbf{x}_0^T D\mathbf{u}\right\} = S.$$

Hence, $\max_{\mathbf{v}} \min_{\mathbf{u}} J(\mathbf{u}, \mathbf{v}) < \min_{\mathbf{u}} \max_{\mathbf{v}} J(\mathbf{u}, \mathbf{v})$, which means that there is no saddle point in pure strategies.

6.5.3 Mixed-Strategy Saddle Point (MSSP)

Since a pure-strategy saddle point PSSP does not exist, we extend the strategy spaces of the players to include randomization (i.e., probability distributions over their action sets). A mixed strategy for the network designer is a probability distribution μ on U_1, and we denote the space of all such probability distributions by M_1. Similarly, a mixed strategy for the adversary is a probability distribution ν on U_2, and the space of all such probability distributions is denoted by M_2. The average cost corresponding to a pair $(\mu \in M_1, \nu \in M_2)$ is given by

$$\bar{J}(\mu, \nu) = \int_{U_1 \times U_2} J(\mathbf{u}, \mathbf{v}) d\mu(\mathbf{u}) d\nu(\mathbf{v}).$$

Definition 2. A pair $(\mu^* \in M_1, \nu^* \in M_2)$ is a mixed-strategy saddle point (MSSP) if the following holds:

$$\bar{J}(\mu^*, \nu) \le \bar{J}(\mu^*, \nu^*) \le \bar{J}(\mu, \nu^*), \text{ for all } (\mu \in M_1, \nu \in M_2).$$

Proposition 2 *Consider the following strategies:*

$$\mu^*(\mathbf{u}) : \mathbf{u} = D^+\mathbf{x}_0 \text{ with probability 1,} \tag{6.24}$$

and

$$\nu^*(\mathbf{v}) : \begin{cases} \mathbf{v} = C\mathbf{p} & \text{with probability 1/2} \\ \mathbf{v} = -C\mathbf{p} & \text{with probability 1/2,} \end{cases} \tag{6.25}$$

where $\mathbf{p}$ *is any unit eigenvector of the matrix* $D^T D$ *corresponding to the largest eigenvalue of* $D^T D$*, that is* $\lambda_{\max}(D^T D)$*. Then the pair* (μ^*, ν^*) *is an MSSP.*

Proof. The cost function is the following:

$$\begin{aligned} J(\mathbf{u}, \mathbf{v}) &= \mathbf{x}_0^T\mathbf{x}_0 + \mathbf{u}^T D^T D\mathbf{u} - 2\mathbf{x}_0^T D\mathbf{u} \\ &\quad + \mathbf{v}^T D^T D\mathbf{v} + 2\mathbf{v}^T \left(D^T D\mathbf{u} - D^T\mathbf{x}_0\right) \\ &= ||D\mathbf{u} - \mathbf{x}_0||_2^2 + \mathbf{v}^T D^T D\mathbf{v} + 2\mathbf{v}^T \left(D^T D\mathbf{u} - D^T\mathbf{x}_0\right). \end{aligned}$$

Then the average cost under the given pair of strategies is,

$$\begin{aligned}\bar{J}(\mu^*, \nu^*) &= ||DD^+\mathbf{x}_0 - \mathbf{x}_0||_2^2 + (C\mathbf{p})^T D^T D(C\mathbf{p}) \times (1/2) \\ &\quad + (-C\mathbf{p})^T D^T D(-C\mathbf{p}) \times (1/2) \\ &= S + C^2 \lambda_{\max}. \end{aligned} \tag{6.26}$$

But we have,

$$\begin{aligned}\bar{J}(\mu^*, \nu) &= ||DD^+\mathbf{x}_0 - \mathbf{x}_0||_2^2 + \int_{U_2} \nu^T D^T D \nu \, \mathrm{d}\nu(\mathbf{v}) \\ &\le S + \max_{\mathbf{v}, ||\mathbf{v}|| \le C} \mathbf{v}^T D^T D \mathbf{v} \\ &= S + C^2 \lambda_{\max} = \bar{J}(\mu^*, \nu^*), \end{aligned} \tag{6.27}$$

$$\begin{aligned}\bar{J}(\mu, \nu^*) &= C^2 \lambda_{\max} + \int_{U_1} ||D\mu - \mathbf{x}_0||_2^2 \, \mathrm{d}\mu(\mathbf{u}) \\ &\ge C^2 \lambda_{\max} + \min_{\mathbf{u}} ||D\mathbf{u} - \mathbf{x}_0||_2^2 \\ &= S + C^2 \lambda_{\max} = \bar{J}(\mu^*, \nu^*). \end{aligned} \tag{6.28}$$

Since we have for any pair $(\mu \in M_1, \nu \in M_2)$,

$$\bar{J}(\mu^*, \nu) \le \bar{J}(\mu^*, \nu^*) \le \bar{J}(\mu, \nu^*),$$

it follows that (μ^*, ν^*) is a saddle-point equilibrium. □

Remark. The saddle point is not unique, as any (μ, ν) where μ is a point distribution in the set $\mathbf{u}_S$ of (6.15) [or any distribution on this set due to the ordered interchangeability property of saddle points (Başar and Olsder 1999)], and ν as in (6.25) where $\mathbf{p}$ is any eigenvector corresponding to $\lambda_{\max}(D^T D)$ (or any distribution on these vectors) is also a saddle point. However, if D is full column rank, and $\lambda_{\max}$ has geometric multiplicity of 1, then the saddle point is unique. □

6.6 Distributed Implementation of the Optimal Control

The computation of an optimal control $\mathbf{u}^* \in \mathbf{u}_S$ (that satisfies the normal equation (6.17) and drives the system to consensus in one iteration) depends on initial values $\mathbf{x}_0$ and the network structure given by the matrix Q. Therefore, it is necessary that the centralized unit be aware of these values to implement this optimal control. However, distributed averaging can run in a decentralized manner by nodes applying the system equation (6.1) in a distributed way:

$$x_i(k+1) = w_{ii}(k)x_i(k) + \sum_{j\in\mathcal{N}_i} w_{ij}(k)x_j(k) \quad \text{for } i = 1,\ldots,n, \tag{6.29}$$

where $\mathcal{N}_i$ is the set of neighbors of node i. Notice that due to Proposition 1, if $W^* = I_n - Q\text{diag}(\mathbf{u}^*)Q^T$, then we have convergence in one iteration:

$$x_{ave} = w_{ii}^* x_i(0) + \sum_{j\in\mathcal{N}_i} w_{ij}^* x_j(0) \quad \text{for } i = 1,\ldots,n.$$

In this section, we study the implementation of the optimal weights W^* in a distributed way. Since $\mathbf{u}^*$ is a minimizer of the convex function $f(\mathbf{u}) = ||D\mathbf{u} - \mathbf{x}_0||_2^2$ where $D = Q\text{diag}(Q^T\mathbf{x}_0)$, then any local minimizer of this function is eventually a global one. We can thus apply the gradient method in a distributed way to lead to convergence to an optimal solution. Our starting point is the following lemma.

Lemma 1. *For any link $l \sim (ij)$ we have*

$$g_l = \frac{\partial f(\mathbf{u})}{\partial u_l} = -2(x_i - x_j)(x_i^+ - x_j^+), \tag{6.30}$$

where $\mathbf{x}^+ = (I - Q\text{diag}(\mathbf{u})Q^T)\mathbf{x}_0 = W\mathbf{x}_0$ and x_i^+ for $i = 1,\ldots,n$ are its elements.

Proof. Since $f(\mathbf{u}) = ||D\mathbf{u} - \mathbf{x}_0||_2^2$,

$$\begin{aligned} g_l &= \frac{\partial ||D\mathbf{u} - \mathbf{x}_0||_2^2}{\partial u_l} \\ &= \left(2D^T(D\mathbf{u} - \mathbf{x}_0)\right)_l \\ &= 2\left(D^T(Q\text{diag}(Q^T\mathbf{x}_0)\mathbf{u} - \mathbf{x}_0)\right)_l \\ &= 2\left(D^T(Q\text{diag}(\mathbf{u})Q^T\mathbf{x}_0 - \mathbf{x}_0)\right)_l \\ &= -2\left(\text{diag}(Q^T\mathbf{x}_0)Q^TW\mathbf{x}_0\right)_l \\ &= -2(Q^T\mathbf{x}_0)_l(Q^TW\mathbf{x}_0)_l \\ &= -2(x_i - x_j)(x_i^+ - x_j^+). \end{aligned} \tag{6.31}$$

The last equality is due to the fact that each column l in the incidence matrix Q has only 2 nonzero terms ($+1$ and -1) corresponding to the nodes incident to the link $l \sim (ij)$. □

Since the gradient at every link can be calculated in a distributed way by Lemma 1, the following iterative gradient method can be used to find the optimal control for the weights on links:

$$u_l(k+1) = u_l(k) - \gamma^{(k)} g_l(k), \tag{6.32}$$

where $\gamma^{(k)}$ is the stepsize. The initial condition can be chosen arbitrarily, for example one choice could be to select $u_l(0) = 0$ for $l = 1, \ldots, m$. Different choices of the stepsize can guarantee the convergence to the optimal solution. For instance, a constant small enough step ($\gamma^{(k)} = \gamma$ for all k) is sufficient for the convergence of the system to the optimal value because the convex function f is differentiable (Polyak 1987, Theorem 1, p. 21), which guarantees that

$$\lim_{k \to \infty} u_l(k) = u_l^*. \tag{6.33}$$

In particular, since the gradient of the $f(\mathbf{u})$ satisfies a Lipschitz condition:

$$\begin{aligned} ||\nabla f(\mathbf{u}_1) - \nabla f(\mathbf{u}_1)|| &= ||2D^T(D\mathbf{u}_1 - \mathbf{x}_0) - 2D^T(D\mathbf{u}_2 - \mathbf{x}_0)|| \\ &= ||2D^T D(\mathbf{u}_1 - \mathbf{u}_2)|| \\ &\leq 2\rho(D^T D)||\mathbf{u}_1 - \mathbf{u}_2|| \\ &= L||\mathbf{u}_1 - \mathbf{u}_2||, \end{aligned}$$

where $\rho(D^T D)$ is the largest eigenvalue of $D^T D$ and $L = 2\rho(D^T D)$ is the Lipschitz constant. Then any constant stepsize such that

$$0 < \gamma < \frac{2}{L} \tag{6.34}$$

would lead to the result in Eq. (6.33).

As a result, letting $W(k) = I - Q\text{diag}(\mathbf{u}(k))Q^T$, the following system

$$\mathbf{y}(k+1) = W(k)\mathbf{x}_0 \tag{6.35}$$

converges to the average, i.e., $\lim_{k\to\infty} \mathbf{y}(k) = \bar{\mathbf{x}}$. Notice that Eq. (6.35) differs from the typical average consensus dynamics in Eq. (6.1) in that $\mathbf{x}_0$ is not changing with time, but it is constant along the iterations while the weight matrix $W(k)$ converges to the optimal value W^* and guarantees the convergence of the system to the average. For a distributed implementation, we have that $g_l(k) = -2(x_i - x_j)(x_i^+ - x_j^+) = -2(x_i - x_j)(y_i(k) - y_j(k))$ for all $l \sim (ij)$. Therefore, at every iteration k, each node first broadcasts its estimate $y_i(k)$ to its neighbors. From the received estimates, each node then calculates $g_l(k)$ using Eq. (6.30) and $u_l(k)$ using Eq. (6.32) for all the links it is incident to. Then, from Eq. (6.35), $y_i(k+1) = w_{ii}(k)x_i(0) + \sum_{j \in \mathcal{N}_i} w_{ij}(k)x_j(0)$ and a new iteration starts.

6.7 Simulations

6.7.1 Adversarial Intervention

We study by simulations the effect of an adversary disrupting the communication on networks having connected random geometric graphs (RGGs) topology. In RGGs, n nodes are thrown uniformly at random on a unit square, and any two nodes within a connectivity radius r are connected by a link. The simulations are done here with a connectivity radius $r = \sqrt{0.6 \times \frac{\log_e(n)}{n}}$ given that the graph is connected. RGGs are generally used as models for wireless sensor networks, and the disruption of communication can be achieved by insertion of high intensity signals on communication links. The additive white noise can also be considered as an adversarial input in our settings. We compare the results on different RGGs with different sizes (number of nodes n) for $n \in \{20, 40, 60, 80, 100\}$. Figure 6.1 shows the different costs on the resulting network with and without the presence of the adversary, averaged over 150 independent runs to achieve 95 % confidence intervals. We consider only one-stage games where the initial cost function is given by $J_0 = \mathbf{x}_0^T \mathbf{x}_0$. For any node i, the initial node value $x_i(0)$ is selected at random uniformly within the interval [0, 1]. We assume that the adversary power constraint is $||\mathbf{v}|| \leq 1$ (i.e., $C = 1$). As shown in Fig. 6.1, as to be expected the network

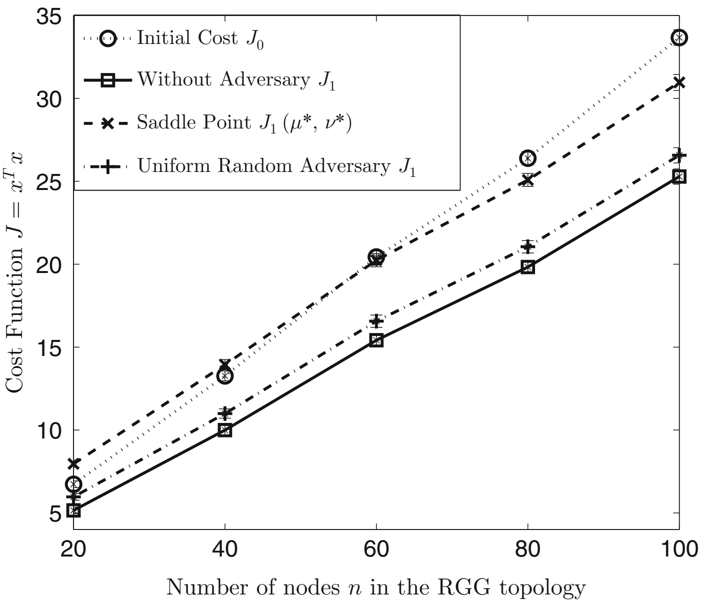

Fig. 6.1 The cost function due to different adversary settings: absence of adversary, uniform random adversary that adds a random noise to the control values, and saddle-point adversary that randomizes its strategy in accordance with the saddle-point equilibrium

without an adversary achieves the least cost J_1. An adversary selecting uniformly random strategy from the n-dimensional unit sphere does not substantially affect the cost; however, an adversary with the same power constraint playing the strategy of the saddle-point equilibrium (Eq. (6.25)) achieves significantly higher cost than the uniform random adversary (even larger cost than J_0 for graphs of $n = 20$ and $n = 40$ nodes). Moreover, we can study the effect of the size of the graph on the cost function. All curves seem to be linear in n. The initial cost J_0 curve shows a higher slope than other cost curves. In fact we can characterize the slope in this case. Since the initial values are selected from a uniform random distribution in the interval [0, 1], we have $\mathbb{E}[J_0] = \mathbb{E}[\mathbf{x}_0^T \mathbf{x}_0] = \sum_i \mathbb{E}[x_i(0)^2] = n\mathbb{E}[X^2] = \frac{1}{3}n$ where X is a uniform random variable lying in the interval [0, 1]. Therefore, the slope of the line J_0 in expectation is $1/3$. Without an adversary, the curve J_1 from simulations shows that it has the smallest slope. Let us characterize the slope in this case. With the given initial conditions, Proposition 1 is satisfied almost surely. Then the network reaches consensus in one iteration, i.e., $\mathbf{x}_1 = x_{ave}\mathbf{1}$. With initial values following a uniform variable between [0, 1], $\mathbb{E}[x_{ave}] = 0.5$ independent of the size of the network. Therefore, $\mathbb{E}[J_1] = \mathbb{E}[\mathbf{x}_1^T \mathbf{x}_1] = nx_{ave}^2 = \frac{1}{4}n$. Hence, the graph without an adversary shows, on the average, a slope of $1/4$ (compared to $1/3$ for the initial distribution). For the graphs with an adversary, the analytic values of the average cost depends not only on the distribution of values, but also on the random graph topology. Based on simulations, the figure shows that the cost due to a uniform random adversary does not change the slope of J_1 graph without the adversary, but only causes some offset. However, these simulations suggest that an adversary with the saddle-point strategy seems to cause a slope slightly larger than $1/4$, but still less than $1/3$. This confirms that an "intelligent" adversary can cause serious harm on the system.

6.7.2 Optimal Control Using Gradient Iterations

As demonstrated in Sect. 6.6, gradient methods can be used to obtain the optimal control for the weights on links. In this part of the simulations, we consider a line network with four nodes, as depicted in Fig. 6.2, where the initial values are picked as $\mathbf{x}_0 = (1, 5, 4, 10)^T$ and the control is given by $\mathbf{u} = (u_{12}, u_{23}, u_{34})^T$. Note that for the given initial values $\mathbf{x}_0$, Proposition 1 says that there exists an optimal control $\mathbf{u}^*$ such that convergence is achieved in one iteration. By applying Eq. (6.18), we obtain that

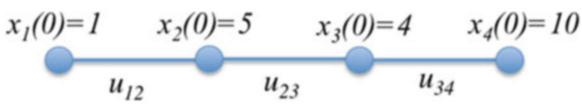

Fig. 6.2 A line network of agents

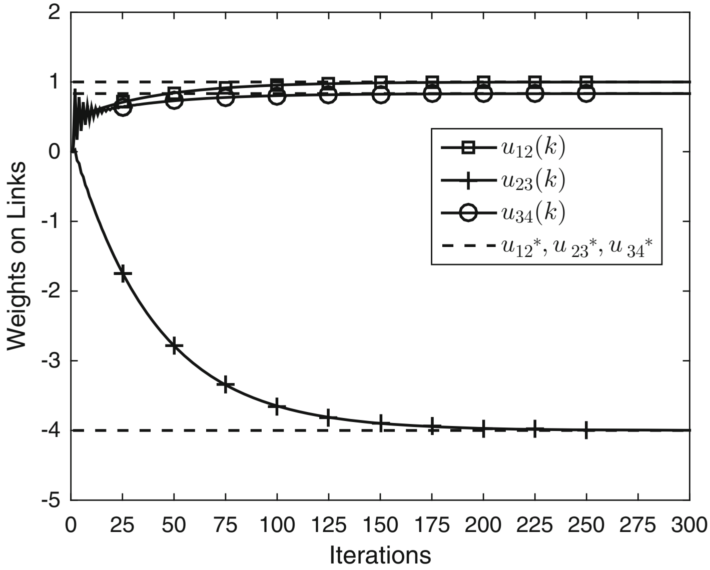

Fig. 6.3 Convergence of weights in the line network using the gradient method

$$\mathbf{u}^* = \left(1, -4, \frac{5}{6}\right)^T.$$

We will apply gradient methods with a constant stepsize $\gamma = 0.0125$ for convergence to the optimal value. Figure 6.3 shows the convergence of the weights on links when using the gradient optimization in Eq. (6.32). The initial starting value for the gradient is $\mathbf{u}(0) = (0, 0, 0)^T$. Figure 6.3 shows that the control variables converge to the optimal values with a reasonable accuracy within 200 iterations.

For further investigation of the proposed gradient method, we consider larger graphs. We do the simulation on RGG graphs with 100 nodes and connectivity radius $r = \sqrt{0.6 \times \frac{\log(n)}{n}} \approx 0.1662$. To measure the distance from consensus at every iteration k, we introduce the metric $e(k)$ defined as follows:

$$e(k) = \log_{10}\left(\frac{L_k}{L_0}\right) = \log_{10}\left(\frac{||\mathbf{x}_k - \bar{\mathbf{x}}||_2^2}{||\mathbf{x}_0 - \bar{\mathbf{x}}||_2^2}\right).$$

For example, $e(k) = -3$ indicates that the current error L_k is 0.1 % of the initial one. This can be used for example as a stopping criterion. Figure 6.4 shows that on RGG graphs, the error from consensus decreases and reaches around 0.1 % with less than 500 iterations. The figure shows the 95 % confidence interval and the resulting graph is an average of 150 independent simulation runs. At each run, each node initial value follows a random variable uniformly distributed between 0 and 1. Note that with these initial values, G' is connected almost surely and thus the optimal control drives the system to consensus in one iteration.

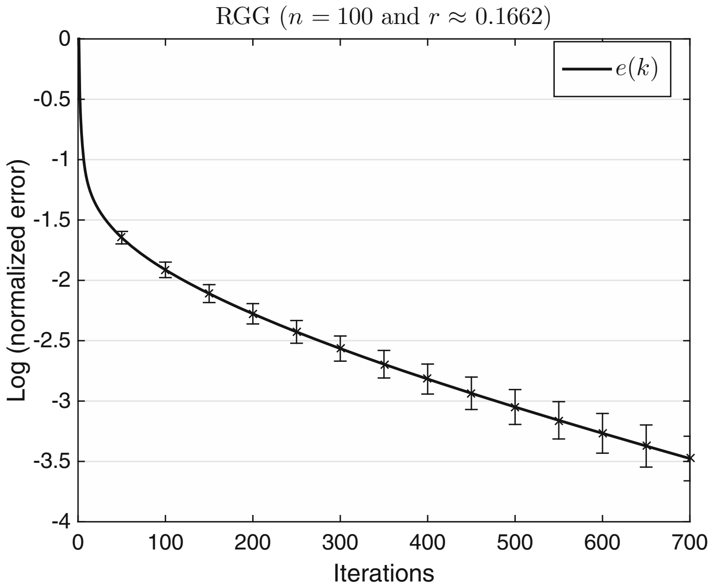

Fig. 6.4 The decrease in the error $e(k)$ with the gradient iterations using a constant stepsize $\gamma = \frac{1}{1.1\rho}$ [to satisfy Eq. (6.34)] shows that the system is eventually converging to consensus (i.e., the gradient method converges to the optimal weights on links)

6.8 Conclusion

In this article, we have studied a zero-sum game between a network designer applying consensus protocols, and an adversary that interrupts these protocols by adding some noise to the weights on communication links. We have studied the saddle-point equilibrium of the consensus problem. We have found that a saddle point in pure strategies does not exist, but it does in mixed strategies. We have obtained the expressions of the mixed strategies where the adversary selects the noise using randomization, whereas the network designer's strategy remains still pure. We have used gradient methods that lead to convergence to this pure strategy in a distributed way. Simulations on random geometric graphs have shown that an adversary adhering to his saddle-point mixed strategy can cause more harm on the system than applying just a random strategy.

For future work, it would be interesting to study the case when the adversary does not have access to initial values, which would then lead to a game with asymmetric information. Studying the equilibrium in the presence of a broader class of adversaries (as malicious and misbehaving nodes, or adversaries that break links) is also one of our future research interests.

Acknowledgements The authors would like to thank Giovanni Neglia for his valuable comments and suggestions to improve the quality of the article. We also acknowledge the U.S. Air Force Office of Scientific Research (AFOSR) MURI grant FA9550-10-1-0573 for partial support of this research.

References

Acemoglu D, Como G, Fagnani F, Ozdaglar A (2011) Opinion fluctuations and persistent disagreement in social networks. In: Proceedings of the 50th IEEE conference on decision and control and European control conference (CDC-ECC 2011), pp 2347–2352

Avrachenkov K, El Chamie M, Neglia G (2011) A local average consensus algorithm for wireless sensor networks. In: Proceedings of the 2011 international conference on distributed computing in sensor systems and workshops (DCOSS), pp 1–6

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. Classics in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia

Ben-Ameur W, Bianchi P, Jakubowicz J (2012) Robust average consensus using total variation gossip algorithm. In: Proceedings of the 6th international conference on performance evaluation methodologies and tools (VALUETOOLS 2012)

Cybenko G (1989) Dynamic load balancing for distributed memory multiprocessors. J Parallel Distrib Comput 7(2):279–301

El Chamie M (2014) Optimization, control, and game theoretical problems in consensus protocols. PhD thesis, Université Nice Sophia Antipolis. https://hal.inria.fr/tel-01103939

El Chamie M, Başar T (2014) Optimal strategies for dynamic weight selection in consensus protocols in the presence of an adversary. In: Proceedings of 53rd IEEE conference on decision and control CDC 2014, Los Angeles, CA, 15–17 December, pp 735–740

El Chamie M, Neglia G, Avrachenkov K (2013) Reducing communication overhead for average consensus. In: Proceedings of IFIP networking 2013, Brooklyn, NY, 22–24 May

El Chamie M, Liu J, Başar T (2014) Design and analysis of distributed averaging with quantized communication. In: Proceedings of the 53rd IEEE conference on decision and control CDC 2014, Los Angeles, CA, 15–17 December, pp 3860–3865

El Chamie M, Neglia G, Avrachenkov K (2015) Distributed weight selection in consensus protocols by Schatten norm minimization. IEEE Trans Autom Control 60(5):1350–1355

Fax J, Murray R (2004) Information flow and cooperative control of vehicle formations. IEEE Trans Autom Control 49(9):1465–1476

Hendrickx JM, Jungers RM, Olshevsky A, Vankeerberghen G (2014) Graph diameter, eigenvalues, and minimum-time consensus. Automatica 50(2):635–640

Hildebrandt TH (1908) Existence of a minimum of a quadratic function. Am Math Mon 15(3):57–59

Khanafer A, Touri B, Başar T (2013) Robust distributed averaging on networks with adversarial intervention. In: Proceedings of the 52nd IEEE conference on decision and control CDC 2013, Florence, 10–13 December, pp 7131–7136

Ko CK, Gao X (2009) On matrix factorization and finite-time average-consensus. In: Proceedings of IEEE CDC/CCC 2009

Lewis FL, Vrabie D, Syrmos VL (2012) Optimal control, 3rd edn. Wiley, New York

Liu Y, Ning P, Reiter MK (2009) False data injection attacks against state estimation in electric power grids. In: Proceedings of the 16th ACM conference on computer and communications security, ACM, New York, CCS '09, pp 21–32

Nedic A, Olshevsky A, Ozdaglar A, Tsitsiklis JN (2009) On distributed averaging algorithms and quantization effects. IEEE Trans Autom Control 54(11):2506–2517

Olfati-Saber R, Murray R (2004) Consensus problems in networks of agents with switching topology and time-delays. IEEE Trans Autom Control 49(9):1520–1533

Olfati-Saber R, Fax J, Murray R (2007) Consensus and cooperation in networked multi-agent systems. Proc IEEE 95(1):215–233

Polyak B (1987) Introduction to optimization. Optimization Software, New York

Ren W, Beard R (2005) Consensus seeking in multiagent systems under dynamically changing interaction topologies. IEEE Trans Autom Control 50(5):655–661

Schwarz V, Matz G (2012) Mean-square optimal weight design for average consensus. In: Proceedings of the 2012 IEEE SPAWC

Vukovic O, Dan G (2014) Security of fully distributed power system state estimation: detection and mitigation of data integrity attacks. IEEE J Sel Areas Commun 32(7):1500–1508

Xiao L, Boyd S (2004) Fast linear iterations for distributed averaging. Syst Control Lett 53:65–78

Chapter 7
Maximal Stable Bridge in Game with Simple Motions in the Plane

Liudmila Kamneva and Valerii Patsko

Abstract It is known that the solvability set (the maximal stable bridge) in a zero-sum differential game with simple motions, fixed terminal time, geometrical constraints for controls of the first and second players, and convex terminal set can be constructed by a program absorption operator. A backward procedure for construction of a t-section of the solvability set does not need any partition of the time segment. In the article, we assert the same property for a game with simple motions, polygonal terminal set (generally non-convex), and polygonal constraints for controls of the players in the plane. In the specific case of a convex terminal set, the operator used in the article coincides with the program absorption operator.

Keywords Differential games with simple motions in the plane • Solvability set • Backward procedure

Math Subject Classifications: 49N70, 49L99, 49N35

7.1 Introduction

In numerical solution of zero-sum differential game, a backward procedure is often used (Subbotin and Patsko 1984; Taras'ev et al. 1987; Sethian 1999; Kumkov et al. 2005; Cristiani and Falcone 2006; Botkin et al. 2011; Dvurechensky and Ivanov 2014) to construct level sets (Lebesgue sets) of the value function. As a rule, at each step $[t_{j-1}, t_j]$, $t_{j-1} < t_j$, of the backward procedure, dynamics of the game is replaced (locally or in global state space) by dynamics of the type $\dot{x} = u + v$, $u \in \mathscr{P}$, $v \in \mathscr{Q}$. Here, x is a state vector; u and v are controls of the first and second

L. Kamneva (✉) • V. Patsko
Krasovskii Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences, S. Kovalevskaya st. 16, Ekaterinburg 620990, Russia

Institute of Mathematics and Computer Science of Ural Federal University, Lenin av. 51, 620083 Ekaterinburg, Russia
e-mail: kamneva@imm.uran.ru; patsko@imm.uran.ru

© Springer International Publishing Switzerland 2016

F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_7

players; $\mathscr{P}$, $\mathscr{Q}$ are convex compact sets. Such a dynamics is called *dynamics of simple motions* (Isaacs 1965). If dimension of x is equal to 2 (i.e. the original game or the equivalent one takes place in the plane) and the game dynamics is replaced globally (it is typical for linear differential games with fixed terminal time), then the level set $\mathscr{W}_c(t_j)$, which corresponds to the value c of the value function at t_j, is approximated by a polygon $W_c(t_j)$, and the compacts $\mathscr{P}$, $\mathscr{Q}$ are substituted with convex polygons P, Q. As a result of the backward procedure, we get a polygon $W_c(t_{j-1})$ approximating the "true" level set $\mathscr{W}_c(t_{j-1})$ of the value function at t_{j-1}.

A transition from $W_c(t_j)$ to $W_c(t_{j-1})$ is often realized by the "program absorption" operator (Krasovskii and Subbotin 1974, p. 122). If the set $W_c(t_j)$ is convex, then the set $W_c(t_{j-1})$ obtained in this case coincides with the exact solution of the differential game of attainability of the set $W_c(t_j)$ for dynamics of simple motions. We have an analogous situation if the complement $\mathbb{R}^2 \setminus W_c(t_j)$, and hence its closure $W'_c(t_j) = \overline{\mathbb{R}^2 \setminus W_c(t_j)}$, is convex. Thus, in the convex case, there exists an operator (namely "program absorption" operator) that gives the exact solution of the approximating problem without any additional partition of the interval $[t_{j-1}, t_j]$.

A natural question is on existence of an operator with the same property, which gives the exact solution $W_c(t_{j-1})$ based on $W_c(t_j)$ in the case when neither the set $W_c(t_j)$ nor its complement is convex.

Our article is dedicated to this question. It is shown that in the case of dynamics of simple motions, an arbitrary polygon $M \subset \mathbb{R}^2$ given at the instant ϑ, and arbitrary convex polygonal constraints P, Q on the controls of the first and second players, there exists an instant $t_* < \vartheta$ such that construction of the set $W_c(t)$, $t \in [t_*, \vartheta)$, can be realized exactly without any additional partition of the interval $[t, \vartheta]$ into smaller subintervals and using them as elements of the backward procedure.

The operator proposed in the article does not coincide, in general, with the "program absorption" operator, but uses it as a "fragment" in a more complex structure. We use a geometric approach to define the operator.

At the end of the article, we give an example of a differential game, for which applying the "program absorption" operator for the original terminal set and its complement gives approximations of the required set from above and below correspondingly. Each of the approximation sets is not coincide precisely with the section of the true solution.

7.2 Differential Game with Simple Motions

Consider a control system with simple motions (Isaacs 1965) in the plane:

$$\dot{x} = u + v, \quad u \in P, \quad v \in Q, \quad t \in [0, \vartheta], \quad \vartheta > 0. \tag{7.1}$$

Here, $x \in \mathbb{R}^2$ is a state vector, u and v are controls of the first and second players, each of the sets P and Q is either a convex closed polygon or a linear segment. (We mean a polygon is a bounded closed set bounded by a polyline without self-intersections and with a finite number of vertices.)

Let M be a given polygon. A differential game is formed by a problem of M-attainability for the first player and a problem of M'-attainability for the second player, $M' = \overline{\mathbb{R}^2 \setminus M}$.

Statement of a problem of M-attainability for the first player is presented in Subbotin (1995, § 13.1) as follows. The first player tries to guarantee $x(\vartheta) \in M$. It is assumed that the player knows the current position $(t, x(t))$ and generates a feedback control $u(t, x(t)) \in P$. To solve the problem of M-attainability, a notion of u-stable bridge is used.

A set-valued function $[0, \vartheta] \ni t \mapsto W(t) \subset \mathbb{R}^2$ defines the *u-stable bridge* (the graph of the function) $W = \{(t, x) : t \in [0, \vartheta],\ x \in W(t)\}$ in the problem of M-attainability if $W(\vartheta) \subset M$, the set W is closed on $[0, \vartheta] \times \mathbb{R}^2$, and for any $v \in Q$ the set W is weakly invariant with respect to the differential inclusion

$$\dot{x} \in P + v. \tag{7.2}$$

The condition of weak invariance means that for any $(t_0, x_0) \in W$ there exists a motion $x(\cdot) : [t_0, \vartheta] \to \mathbb{R}^2$ satisfying differential inclusion (7.2), the initial condition $x(t_0) = x_0$ and the viability condition: $x(t) \in W(t)$ for all $t \in [t_0, \vartheta]$. In the theory of differential game, this property (in an equivalent formulation) is called the u-stability condition.

In the same way, the problem of M'-attainability for the second player is formulated and the notion of v-stable bridge is introduced.

The original (equivalent) notions of stable bridges were presented in Krasovskii and Subbotin (1974, pp. 52–54) and Krasovskii and Subbotin (1988, pp. 53, 58).

Let W_0 denote *the maximal* (*by inclusion*) *u-stable bridge* in the problem of M-attainability. In Cardaliaguet et al. (1999), an analogues set is called *a discriminating kernel*.

It is known (Krasovskii and Subbotin 1974, § 16) that the set W is the maximal u-stable bridge in the problem of M-attainability if and only if the set $W' = \{(t, x) : t \in [0, \vartheta],\ x \in \overline{\mathbb{R}^2 \setminus W(t)}\}$ is the maximal v-stable bridge in the problem of M'-attainability. Thus, if we construct the maximal u-stable bridge, we get both the solutions of M-attainability and M'-attainability problems.

The following property is true. Assume that for any $t \in [t_1, t_2]$ the section $W_0(t)$ of the maximal stable bridge W_0 is a polygon. Then the set-valued function $[t_1, t_2] \ni t \to W_0(t)$ is continuous with respect to the Hausdorff metric topology.

The article aims to investigate a possibility of constructing the maximal stable bridge for some interval $[t_*, \vartheta]$ by an operator, which does not need any additional partition of the interval.

7.3 Spirals in the Plane, Semipermeable Tubes, and Surfaces

7.3.1 Spiral Polylines in the Plane

Polyline $\gamma = a_0 a_1 \dots a_{n_\gamma}$ in the plane is a union of a finite number of linear segments $[a_i, a_{i+1}]$, $i = \overline{0, n_\gamma - 1}$, $a_i \neq a_{i+1}$, such that an end of each segment (except possibly the last one) is the beginning of the next segment while segments that have a common end do not belong to a straight line. A segment is also considered as a polyline. Points $a_0, a_1, \dots, a_{n_\gamma}$ are called *vertices* of the polyline; segments $[a_i, a_{i+1}]$, $i = \overline{0, n_\gamma - 1}$, are called *edges* of the polyline; segments that have a common vertex are called *adjacent* edges. A polyline γ is *closed* if the end of the last edge coincides with the beginning of the first one, i.e. $a_{n_\gamma} = a_0$.

Normally, a polyline γ is called *a convex polyline* if it is in the same half-plane with respect to any straight line containing an edge of the polyline (so, a closed convex polyline bounds a convex polygon). The boundary of a nonconvex polygon can be represented as a union of nonclosed polylines of different types. To do this, we generalize the notion of a nonclosed convex polyline as follows.

Definition 1. *An angle of a polyline γ at a vertex a_i, $i = \overline{1, n_\gamma - 1}$,* is an oriented (with respect to the sign) angle less then π formed by rays with the direction vectors $\overrightarrow{a_i a_{i+1}}$, $\overrightarrow{a_i a_{i-1}}$ and the vertex at the point a_i. *An adjacent angle of the polyline γ at the vertex a_i, $i = \overline{1, n_\gamma - 1}$,* is an oriented angle less then π formed by the rays with the direction vectors $\overrightarrow{a_{i-1} a_i}$, $\overrightarrow{a_i a_{i+1}}$ and the beginning at a_i. [An angle is called positive (negative) if the transition from the first to the second direction vector is counterclockwise (clockwise). The sum of the angle of a polyline at the vertex a_i, $i = \overline{1, n_\gamma - 1}$, and the corresponding adjacent angle is equal to $\pm\pi$.]

Definition 2. A nonclosed polyline without self-intersections is called *convex* if its adjacent angles have the same sign.

Definition 3. *A single left (right) spiral* is defined to be a nonclosed polyline without self-intersections such that all its adjacent angles are positive (negative) and its first edge $[a_0, a_1]$ belongs to the boundary of its convex hull.

Let an edge $[a_i, a_{i+1}]$ of a single spiral γ be assigned to a unit normal vector with direction obtained by rotating the vector $\overrightarrow{a_i a_{i+1}}$ by the angle $-\pi/2$ $(+\pi/2)$ for left (right) spiral, $i = \overline{0, n_\gamma - 1}$. Note that the unit normal vectors for edges from the boundary of the convex hull are directed outward the convex hull (Fig. 7.1a, b). The "outward" and "inward" end vertices of the single spiral γ are denoted by a_γ and b_γ correspondingly.

Definition 4. A polyline γ is called *a double spiral* if it can be represented as a union $\gamma = \gamma_0^l \cup \gamma_0 \cup \gamma_0^r$, where γ_0^l, γ_0, γ_0^r are nonclosed polylines, $\gamma_0 = \gamma \cap \partial(\mathrm{co}\,\gamma)$, $\gamma_0^l \cap \gamma_0^r = \varnothing$, $\gamma_0 \cup \gamma_0^l = \gamma^l$ is a left spiral, and $\gamma_0 \cup \gamma_0^r = \gamma^r$ is a right spiral.

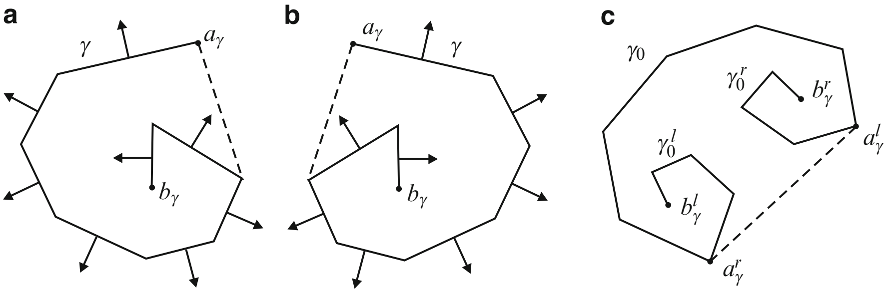

Fig. 7.1 Single left (**a**) and right (**b**) spirals; double spiral (**c**)

Let the end vertices of the spiral γ^l be denoted by a_γ^l, b_γ^l; and the end vertices of the spiral γ^r be denoted by a_γ^r, b_γ^r (Fig. 7.1c).

Changing the numeration of vertices, we can refer any nonclosed convex polyline to one of the following types: a single left spiral, a single right spiral, a double spiral.

7.3.2 Semipermeable Tube

Let $O(z,\varepsilon)$ be an open circle of center z and radius $\varepsilon > 0$ in $\mathbb{R}^2$.

For $\varepsilon > 0$, define a value $\Delta(\varepsilon) > 0$ such that for any $z \in \mathbb{R}^2$ a trajectory of (7.1) with an initial position in the set $O(z,\varepsilon/2)$ does not leave the set $O(z,\varepsilon)$ on the interval $[0,\Delta(\varepsilon)]$.

Let $0 \le t_1 < t_2 \le \vartheta$, and consider a continuous (by Hausdorff metric) set-valued function $[t_1,t_2] \ni t \mapsto W(t) \subset \mathbb{R}^2$ such that $W(t)$ is a polygon for all $t \in [t_1,t_2]$. The mapping $W(\cdot)$ defines the set (graph) $W = \{(t,x) :\ t \in [t_1,t_2],\ x \in W(t)\}$.

We denote the lateral boundary of W by Γ:

$$\Gamma = \{(t,x) :\ t \in [t_1,t_2],\ x \in \partial W(t)\}.$$

The set Γ is called *a tube* on the segment $[t_1,t_2]$. Let $\Gamma(t) = \partial W(t)$ be a section of the tube at t. Define an ε-neighborhood of the tube Γ:

$$O_\varepsilon(\Gamma) = \{(t,x) :\ t \in [t_1,t_2],\ x \in O(z,\varepsilon),\ z \in \Gamma(t)\}.$$

Definition 5. A tube Γ is called *semipermeable* on the segment $[t_1,t_2]$ if there exist $\varepsilon > 0$ and open sets $O_\varepsilon^+(\Gamma)$, $O_\varepsilon^-(\Gamma)$ such that

$$O_\varepsilon(\Gamma) = O_\varepsilon^+(\Gamma) \cup \Gamma \cup O_\varepsilon^-(\Gamma), \quad O_\varepsilon^+(\Gamma) \cap O_\varepsilon^-(\Gamma) = \varnothing,$$

and

1) for any $(t_0, x_0) \in O_{\varepsilon/2}(\Gamma) \cap \big(O_\varepsilon^+(\Gamma) \cup \Gamma\big)$ and $v \in Q$ there exists a measurable open-loop control $u(t) \in P$ such that the solution $x(t)$ to the equation $\dot{x}(t) = u(t) + v$ with an initial condition $x(t_0) = x_0$ satisfies the inclusion $(t, x(t)) \in O_\varepsilon^+(\Gamma) \cup \Gamma$ for all $t \in [t_0, t_0 + \Delta(\varepsilon)] \cap [t_1, t_2]$;
2) for any $(t_0, x_0) \in O_{\varepsilon/2}(\Gamma) \cap \big(O_\varepsilon^-(\Gamma) \cup \Gamma\big)$ and $u \in P$ there exists a measurable open-loop control $v(t) \in Q$ such that the solution $x(t)$ to the equation $\dot{x}(t) = u + v(t)$ with an initial condition $x(t_0) = x_0$ satisfies the inclusion $(t, x(t)) \in O_\varepsilon^-(\Gamma) \cup \Gamma$ for all $t \in [t_0, t_0 + \Delta(\varepsilon)] \cap [t_1, t_2]$.

Given the definition, we use notions of side $(+)$ and side $(-)$ of the semipermeable tube. We distinguish two types of semipermeable tubes: if the side $(+)$ is internal and the side $(-)$ is external, then the semipermeable tube has a type $\pm$; otherwise, the semipermeable tube has a type $\mp$.

We have that if the lateral surface $\Gamma_0 = \{(t, x) : \ t \in [0, \vartheta], \ x \in \partial W_0(t)\}$ of the maximal u-stable bridge W_0 is a tube, then Γ_0 is a semipermeable tube of type $\pm$. The results of the books (Isaacs 1965; Krasovskii and Subbotin 1974, 1988) imply the converse:

Lemma 1. *Assume that Γ is a semipermeable tube of type $\pm$ on a segment $[t_1, \vartheta]$, $t_1 \in [0, \vartheta)$, and $\Gamma(\vartheta) = \partial M$. Then Γ is the lateral surface of the maximal u-stable bridge W_0 on $[t_1, \vartheta]$ in the problem of M-attainability.*

7.3.3 Semipermeable Surfaces

Let $0 \le t_1 < t_2 \le \vartheta$, and consider a continuous (by Hausdorff metric) set-valued function $[t_1, t_2] \ni t \mapsto \sigma(t) \subset \mathbb{R}^2$ such that $\sigma(t)$ is a nonclosed polyline without self-intersections. Additionally we assume that the end vertices of the polyline $\sigma(t)$ form two continuous trajectories without self-intersections for $t \in [t_1, t_2]$. The mapping $\sigma(\cdot)$ defines the set (graph) $\sigma = \{(t, x) : \ t \in [t_1, t_2], \ x \in \sigma(t)\}$, which is called *a surface*. Let us denote by $\hat{\xi}(\cdot)$ and $\check{\xi}(\cdot)$ the "end trajectories" of the graph, and we will use the notion of a surface σ with edges $\hat{\xi}(\cdot)$ and $\check{\xi}(\cdot)$.

If $\sigma(t)$ is a spiral for all $t \in [t_1, t_2]$, then the surface σ is called *spiral*.

For $\varepsilon > 0$, we define an ε-neighborhood of the surface σ by

$$O_\varepsilon(\sigma) = \{(t, x) : \ t \in [t_1, t_2], \ x \in O(z, \varepsilon), \ z \in \sigma(t)\}.$$

Definition 6. A surface σ is called *semipermeable* on $[t_1, t_2]$ if there exist $\varepsilon > 0$, open sets $O_\varepsilon^+(\Gamma)$, $O_\varepsilon^-(\Gamma)$, and a surface $[\sigma]_\varepsilon$ such that

$$O_\varepsilon(\sigma) = O_\varepsilon^+(\sigma) \cup [\sigma]_\varepsilon \cup O_\varepsilon^-(\sigma), \quad O_\varepsilon^+(\sigma) \cap O_\varepsilon^-(\sigma) = \varnothing, \quad \sigma \subset [\sigma]_\varepsilon, \qquad (7.3)$$

and

1) for any $(t_0, x_0) \in O_{\varepsilon/2}(\sigma) \cap \left(O_\varepsilon^+(\sigma) \cup [\sigma]_\varepsilon\right)$ and $v \in Q$ there exists a measurable open-loop control $u(t) \in P$ such that the solution $x(t)$ to the equation $\dot{x}(t) = u(t) + v$ with an initial condition $x(t_0) = x_0$ satisfies the inclusion $(t, x(t)) \in O_\varepsilon^+ \cup [\sigma]_\varepsilon$ for all $t \in [t_0, t_0 + \Delta(\varepsilon)] \cap [t_1, t_2]$;
2) for any $(t_0, x_0) \in O_{\varepsilon/2}(\sigma) \cap \left(O_\varepsilon^-(\sigma) \cup [\sigma]_\varepsilon\right)$ and $u \in P$ there exists a measurable open-loop control $v(t) \in Q$ such that the solution $x(t)$ to the equation $\dot{x}(t) = u + v(t)$ with an initial condition $x(t_0) = x_0$ satisfies the inclusion $(t, x(t)) \in O_\varepsilon^- \cup [\sigma]_\varepsilon$ for all $t \in [t_0, t_0 + \Delta(\varepsilon)] \cap [t_1, t_2]$.

Given the definition, we use notions of side $(+)$ and side $(-)$ of the semipermeable surface.

We distinguish two types of semipermeable surfaces: if the unit normal vectors of the spiral are directed from side $(+)$ to side $(-)$, then the semipermeable surface has a type $\pm$; otherwise the semipermeable surface is said to be of type $\mp$.

By $S^+(\cdot) = S^+(\,\cdot\,; \Gamma, [t_1, t_2], \hat{\xi}(\cdot), \check{\xi}(\cdot))$ we mean a set-valued function taking a value $t \in [t_1, t_2]$ to a polyline $S^+(t)$ that is a part of the section $\Gamma(t)$ of the tube Γ obtained by moving along $\Gamma(t)$ from $\hat{\xi}(t)$ to $\check{\xi}(t)$ counterclockwise. Analogously we define a set-valued function $S^-(\cdot) = S^-(\cdot; \Gamma, [t_1, t_2], \hat{\xi}(\cdot), \check{\xi}(\cdot))$ moving along $\Gamma(t)$ from $\hat{\xi}(t)$ to $\check{\xi}(t)$ clockwise.

The definitions of a semipermeable tube and a semipermeable surface directly imply the following lemma.

Lemma 2. *Assume that Γ is a semipermeable tube on $[t_1, t_2]$. Then the mapping $\sigma(\cdot) = S^+(\cdot) = S^+(\,\cdot\,; \Gamma, [t_1, t_2], \hat{\xi}(\cdot), \check{\xi}(\cdot))$ defines a semipermeable surface σ with edges $\hat{\xi}(\cdot)$, $\check{\xi}(\cdot)$ for any continuous disjoint trajectories $\hat{\xi}(t)$, $\check{\xi}(t)$, $t \in [t_1, t_2]$, on the tube Γ. The same property is valid for the mapping $S^-(\cdot)$.*

Next, we formulate and prove an assertion on sewing semipermeable surfaces.

Lemma 3. *Assume that*

a) σ_i is a semipermeable surface with edges $\hat{\xi}_i(\cdot)$, $\check{\xi}_i(\cdot)$ on $[t_1, t_2]$, $i = 1, 2$;
b) $\sigma_1 \cap \sigma_2$ is a surface with edges $\hat{\xi}_2(\cdot)$, $\check{\xi}_1(\cdot)$ on $[t_1, t_2]$;
c) $\sigma = \sigma_1 \cup \sigma_2$ is a surface with the edges $\hat{\xi}_1(\cdot)$, $\check{\xi}_2(\cdot)$;
d) the side $(+)$ (side $(-)$) of the surface σ_1 adjoins the side $(+)$ (side $(-)$) of the surface σ_2 via the intersection $\sigma_1 \cap \sigma_2$ forming the side $(+)$ (side $(-)$) of the surface σ.

Then σ is a semipermeable surface with definite sides $(+)$ and $(-)$.

Proof. For $t \in [t_1, t_2]$, we consider the point $\xi^*(t)$ that divides the polyline $\sigma_1(t) \cap \sigma_2(t)$ into two polylines of equal length. The "average" path $\xi^*(\cdot)$ is a continuous line and separates two parts σ_1^* and σ_2^* with edges $\hat{\xi}_1(\cdot)$, $\xi^*(\cdot)$ and $\xi^*(\cdot)$, $\check{\xi}_2(\cdot)$ correspondingly; a schematic view is shown in Fig. 7.2a.

Let h_0 be the minimal distance between the end vertices $\hat{\xi}_2(t)$ and $\check{\xi}_1(t)$ of the polyline $\sigma_1(t) \cap \sigma_2(t)$ for $t \in [t_1, t_2]$, ε_i is a value defined in (7.3) by semipermeable

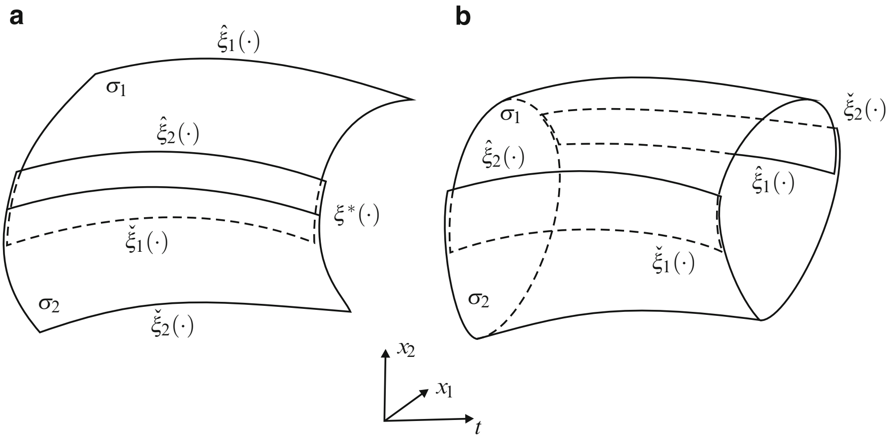

Fig. 7.2 Sewing semipermeable surfaces into a semipermeable surface (**a**); a semipermeable tube (**b**)

property of the surface σ_i, $i = 1, 2$. Fix $\varepsilon \in (0, \min\{\varepsilon_1, \varepsilon_2, h_0/2\})$. We observe that

$$O(\xi^*(t), \varepsilon) \cap \sigma_i(t) = O(\xi^*(t), \varepsilon) \cap \sigma(t), \quad i = 1, 2, \quad t \in [t_1, t_2].$$

Therefore, $O_\varepsilon(\sigma) = O_\varepsilon(\sigma_1^*) \cup O_\varepsilon(\sigma_2^*)$.

Since $O_\varepsilon(\sigma_i^*) \subset O_{\varepsilon_i}(\sigma_i)$, $\sigma_i^* \subset \sigma_i$, we have

$$O_\varepsilon(\sigma_i^*) = O_\varepsilon^+(\sigma_i^*) \cup [\sigma_i^*]_\varepsilon \cup O_\varepsilon^-(\sigma_i^*),$$

where

$$O_\varepsilon^\pm(\sigma_i^*) := O_{\varepsilon_i}^\pm(\sigma_i) \cap O_\varepsilon(\sigma_i^*), \quad [\sigma_i^*]_\varepsilon := [\sigma_i]_{\varepsilon_i} \cap O_\varepsilon(\sigma_i^*), \quad i = 1, 2. \tag{7.4}$$

Define

$$O_\varepsilon^\pm(\sigma) = O_\varepsilon^\pm(\sigma_1^*) \cup O_\varepsilon^\pm(\sigma_2^*), \quad [\sigma]_\varepsilon = [\sigma_1^*]_\varepsilon \cup [\sigma_2^*]_\varepsilon.$$

The sets $O_\varepsilon^\pm(\sigma)$ are open, $O_\varepsilon^+(\sigma) \cap O_\varepsilon^-(\sigma) = \varnothing$, and $\sigma \subset [\sigma]_\varepsilon$.

Thus, we deduce the representation

$$O_\varepsilon(\sigma) = O_\varepsilon^+(\sigma) \cup [\sigma]_\varepsilon \cup O_\varepsilon^-(\sigma).$$

Properties 1) and 2) from the definition of a semipermeable surface are valid by virtue of (7.4).

In the same way, the following lemma can be proved (see Fig. 7.2b).

Lemma 4. *Assume that*

a) σ_i *is a semipermeable surface with edges* $\hat{\xi}_i(\cdot)$, $\check{\xi}_i(\cdot)$ *on* $[t_1, t_2]$, $i = 1, 2$;
b) the intersection $\sigma_1 \cap \sigma_2$ *consists of two disjoint surfaces with edges* $\hat{\xi}_2(\cdot)$, $\check{\xi}_1(\cdot)$ *and* $\hat{\xi}_1(\cdot)$, $\check{\xi}_2(\cdot)$ *correspondingly;*
c) $\sigma = \sigma_1 \cup \sigma_2$ *is a tube;*
d) the side $(+)$ (*side* $(-)$) *of the surface* σ_1 *adjoins the side* $(+)$ (*side* $(-)$) *of the surface* σ_2 *via the intersection* $\sigma_1 \cap \sigma_2$ *forming the side* $(+)$ (*side* $(-)$) *of the tube* σ.

Then σ *is a semipermeable tube with definite sides* $(+)$ *and* $(-)$.

7.4 Program Absorption Operator and Its Effect on Convex Sets

We construct a semipermeable surface by sewing pieces of lateral boundaries of the maximal stable bridges for convex polygonal terminal sets. To this end, we formulate some assertions that are used in our analysis of t-sections of the bridges. The corresponding proofs are given in the Appendix.

Let $\Pi(\alpha, \nu) := \{x \in \mathbb{R}^2 : \langle x, \nu \rangle \leq \alpha\}$ be a half-plane defined by a value $\alpha \in \mathbb{R}$ and a unit vector $\nu \in \mathbb{R}^2$ which is an outer normal to the boundary of the half-plane; $\mathcal{N}(A)$ is the set of outer unit normals to the edges of a convex polygon A; $\rho(\eta; A) = \sup\{\langle a, \eta \rangle : a \in A\}$ is a value of the support function of a set A at a vector η. Then $\Pi(\rho(\eta; A), \eta)$ is a supporting half-plane of the set A with the outer normal η.

For $\tau \in [0, \vartheta]$ and an arbitrary set $A \subset \mathbb{R}^2$, consider

$$T_\tau(A) := \{x \in \mathbb{R}^2 : x + \tau q \in A - \tau P, \quad \forall\, q \in Q\} = \bigcap_{q \in Q} (A - \tau(P + q)).$$

The operator $T_\tau : A \to T_\tau(A)$ is called (Krasovskii and Subbotin 1974) *a program absorption operator.* We use it for $A \subset \mathbb{R}^2$.

Note the following properties of the operator T_τ (Pshenichnyy and Sagaydak 1971; Kamneva and Patsko 2013):

(1) if $A \subset B$, then $T_\tau(A) \subset T_\tau(B)$ for all $\tau > 0$;
(2)

$$T_\tau(A \cap B) \subset T_\tau(A) \cap T_\tau(B); \tag{7.5}$$

(3) if a set A is convex and $\tau_1, \tau_2 > 0$, then $T_{\tau_1}(T_{\tau_2}(A)) = T_{\tau_1 + \tau_2}(A)$;
(4) for a half-plane $\Pi(\alpha, \nu)$, we have

$$T_\tau(\Pi(\alpha, \nu)) = \Pi(\alpha, \nu) - \tau(p_* + q_*),$$

where

$$p_* \in \text{Arg} \min_{p \in P} \langle p, \nu \rangle, \quad q_* \in \text{Arg} \max_{q \in Q} \langle q, \nu \rangle$$

[consequently the set $\{(\tau, x) : \tau > 0, \ x \in \partial T_\tau(\Pi(\alpha, \nu))\}$ is a half-plane of variables (τ, x) in $\mathbb{R} \times \mathbb{R}^2$];

(5) for a half-plane $\Pi(\alpha, \nu)$ and $r \in \mathbb{R}$, we have

$$T_\tau(\Pi(\alpha, \nu) + r\nu) = T_\tau(\Pi(\alpha, \nu)) + r\nu. \tag{7.6}$$

We define dist (A, B) to be the minimal distance between any two points of the sets A and B:

$$\text{dist}\,(A, B) = \inf_{a \in A,\, b \in B} \|a - b\|.$$

Lemma 5. *Assume that A is a convex set in $\mathbb{R}^2$, $\Pi_0 = \Pi(\alpha_0, \nu_0)$ is a half-plane, $A \subset \Pi_0$, $\tau > 0$, and $T_\tau(A) \neq \varnothing$. Then* $\text{dist}\,(T_\tau(A), \partial T_\tau(\Pi_0)) \geq \text{dist}\,(A, \partial \Pi_0)$.

Lemma 6. *Assume that A is a convex polygon, each of the sets P and Q is either a convex polygon or a segment, and $\tau > 0$. Then*

$$T_\tau(A) = \bigcap \{T_\tau(\Pi(\rho(\eta, A), \eta)) : \ \eta \in \mathscr{N}(A) \cup \mathscr{N}(-P)\}.$$

Write

$$\mathscr{R}_P(A) := \{\Pi(\rho(\eta; A), \eta) : \ \eta \in \mathscr{N}(A) \cup \mathscr{N}(-P)\} \tag{7.7}$$

for extended set of supporting half-planes of the set A.

Lemma 7. *Let A and B be convex polygons, and assume $B \subset A$. Then*

$$T_\tau(B) = T_\tau(A) \bigcap \Big(\bigcap \{T_\tau(\Pi) : \ \Pi \in \mathscr{R}_P(B) \setminus \mathscr{R}_P(A)\} \Big). \tag{7.8}$$

Lemma 8. *Let A be a convex polygon, and let Π be a half-plane. Assume that $\tau_0 > 0$, $T_{\tau_0}(A) \subset T_{\tau_0}(\Pi)$, and $T_{\tau_0}(A) \cap \partial T_{\tau_0}(\Pi)$ is an edge of the polygon $T_{\tau_0}(A)$. Then $T_\tau(A) \cap \partial T_\tau(\Pi)$ is an edge of the polygon $T_\tau(A)$ for all $\tau \in (0, \tau_0)$.*

Lemma 9. *Let A and B be convex polygons. Assume $B \subset A$, $\partial A \cap \partial B$ is a nondegenerate nonclosed polyline, $\tau > 0$, and $T_\tau(B) \neq \varnothing$. Then the set $\partial T_\tau(A) \cap \partial T_\tau(B)$ is either an empty set or a nonempty connected set (a point, a closed or nonclosed polyline), and the set-valued mapping $\tau \to \partial T_\tau(A) \cap \partial T_\tau(B)$ is continuous.*

7.5 Construction of Semipermeable Spiral Surfaces

7.5.1 Spiral System $\mathscr{A}^l(\gamma, P)$ of Convex Sets for a Single Left Spiral

Let $\gamma = a_0 a_1 \dots a_{n_\gamma}$ be a single left spiral. Define ν_i to be the unit normal to the edge $[a_{i-1}a_i]$ of the spiral γ, and let $\mathscr{N}_\gamma = \{\nu_1, \nu_2, \dots, \nu_{n_\gamma}\}$. Remind that, for a left spiral, the direction of the normal to the edge $[a_{i-1}a_i]$ is obtained by rotating the vector $\overrightarrow{a_{i-1}a_i}$ by the angle $(-\pi/2)$. The vector ν_i rotates counterclockwise while the subscript increases (for a left spiral).

Let us form a new set of normals $\mathscr{N}_\gamma^P$ by adding the set $\mathscr{N}(-P)$ to the set $\mathscr{N}_\gamma$ as follows. Looking over all pairs of adjacent normals ν_i and ν_{i+1}, $i = \overline{1, n_\gamma - 1}$, we find all the vectors from $\mathscr{N}(-P)$ belonging to an open cone formed by the vectors ν_i and ν_{i+1}, and we arrange them *counterclockwise* (for the left spiral): $\eta_i^1, \eta_i^2, \dots, \eta_i^{\beta_i}$. Such a set may be empty. We insert these sets of vectors between the respective pairs of the normals ν_i and ν_{i+1}, and we denote the result set by $\mathscr{N}_\gamma^P$:

$$\mathscr{N}_\gamma^P = \{\nu_1, \eta_1^1, \dots, \eta_1^{\beta_1}, \nu_2, \eta_2^1, \dots, \eta_2^{\beta_2}, \nu_3, \dots, \nu_{n_\gamma - 1}, \eta_{n_\gamma-1}^1, \dots, \eta_{n_\gamma-1}^{\beta_{n_\gamma-1}}, \nu_{n_\gamma}\}.$$

For any vector $h \in \mathscr{N}_\gamma^P$, consider the half-plane $\Pi^\gamma(h)$:

$$\Pi^\gamma(h) = \{z \in \mathbb{R}^2 : \langle z - a_i, h\rangle \le 0\} \quad \text{if } \; h = \nu_i \text{ or } \; h = \eta_i^j, \quad j = \overline{1, \beta_i}.$$

Here, the index i is defined by vector h.

We introduce a uniform indication of the elements of the set $\mathscr{N}_\gamma^P$:

$$\mathscr{N}_\gamma^P = \{h_1, \dots, h_m\}.$$

Let the spiral γ be associated to *a spiral system*

$$\mathscr{A}^l(\gamma, P) = \{A_1 \supset A_2 \supset A_3 \supset \dots \supset A_n\}$$

of nested convex sets A_i, $i = \overline{1, n}$, as follows. Define $A_1 := \operatorname{co}\gamma$. If $\gamma \subset \partial A_1$, then write $n = 1$ and $\mathscr{A}^l(\gamma, P) = \{A_1\}$. In the case $\gamma \not\subset \partial A_1$, the value $n_1 = \max\{i \in \overline{1, m} : \; A_1 \subset \Pi^\gamma(h_i)\}$ satisfies the inequality $n_1 < m$. Consider auxiliary sets

$$A_{n_1}^* = A_1, \quad A_j^* = A_{j-1}^* \cap \Pi^\gamma(h_j), \quad j = \overline{n_1 + 1, m}.$$

For all $j = \overline{n_1 + 1, m}$, the boundary $\partial\Pi^\gamma(h_j)$ of the half-plane $\Pi^\gamma(h_j)$ divides the set A_{j-1}^* into two non-empty subsets by the segment $[\hat{e}_j^*, \check{e}_j^*] = \partial\Pi^\gamma(h_j) \cap A_{j-1}^*$. We choose the notations for the vertices of the segment in such a way that the half-plane $\Pi^\gamma(h_j)$ is *on the left hand* (for the left spiral) when moving along the segment from

$\hat{e}^*_j$ to $\check{e}^*_j$. Note that

$$A^*_j = A_1 \bigcap \left(\bigcap_{i=n_1+1}^{j} \Pi^\gamma(h_i) \right), \quad j = \overline{n_1+1, m}.$$

We define the set A_2 on the basis of the set A_1 as follows. Let

$$\Phi^j_2 := \bigcap_{i=n_1+1}^{j} \Pi^\gamma(h_i) = \bigcap \{\Pi : \ \Pi \in \mathscr{R}_P(A^*_j) \setminus \mathscr{R}_P(A_1)\}, \quad j = \overline{n_1+1, m}.$$

Note that $\Phi^{n_1+1}_2$ is a half-plane and, consequently, it is unbounded. Find the maximal number $n_2 \in \overline{n_1+1, m}$ such that the set $\Phi^{n_2}_2$ is unbounded.

Write

$$A_2 := A^*_{n_2} = A_1 \bigcap \left(\bigcap_{i=n_1+1}^{n_2} \Pi^\gamma(h_i) \right) = A_1 \cap \Phi^{n_2}_2, \quad \hat{e}_2(\vartheta) := \hat{e}^*_{n_1+1}, \quad \check{e}_2(\vartheta) := \check{e}^*_{n_2}.$$

If $n_2 = m$, then write $n = 2$ and $\mathscr{A}^l(\gamma, P) = \{A_1 \supset A_2\}$. In the case $n_2 < m$, we construct the set A_3 on the basis of the set A_2. And so on.

Generally, let us define the set A_k on the basis of the set A_{k-1} for the case $n_{k-1} < m$. Let

$$\Phi^j_k := \bigcap_{i=n_{k-1}+1}^{j} \Pi^\gamma(h_i) = \bigcap \{\Pi : \ \Pi \in \mathscr{R}_P(A^*_j) \setminus \mathscr{R}_P(A_{k-1})\}, \ j = \overline{n_{k-1}+1, m}. \tag{7.9}$$

The set $\Phi^{n_{k-1}+1}_k$ is a half-plane. Find the maximal number $n_k \in \overline{n_{k-1}+1, m}$ such that the set $\Phi^{n_k}_k$ is unbounded.

Write

$$A_k := A^*_{n_k} = A_{k-1} \bigcap \left(\bigcap_{i=n_{k-1}+1}^{n_k} \Pi^\gamma(h_i) \right) = A_{k-1} \cap \Phi^{n_k}_k,$$

$$\hat{e}_k(\vartheta) := \hat{e}^*_{n_{k-1}+1}, \quad \check{e}_k(\vartheta) := \check{e}^*_{n_k}.$$

If $n_k = m$, we have $n = k$ and $\mathscr{A}^l(\gamma, P) = \{A_1 \supset A_2 \supset \cdots \supset A_k\}$.

Figure 7.3 illustrates a construction of a spiral system. Here, a left spiral γ consists of ten edges. The set P is a triangle. The set $-P$ and its outer unit normals are shown. The normal h_3 to the third edge of the polyline γ is coincide with one of the normal to the triangle $-P$. There are no others coinciding normals. We have $m = 14$.

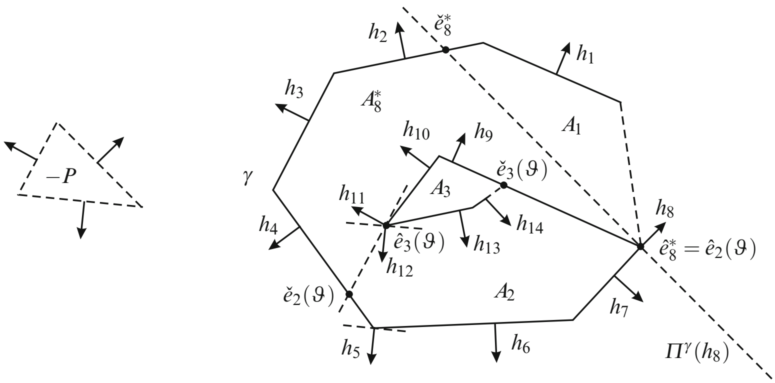

Fig. 7.3 Explanation of construction of a spiral system $\mathscr{A}^l(\gamma, P) = \{A_1 \supset A_2 \supset A_3\}$ for $n_1 = 7$, $n_2 = 11$, and $n_3 = m = 14$

The set A_1 is defined as $A_1 = \operatorname{co}\gamma$. Then, starting with the normal h_1, we find a normal with the maximal number n_1 in the set h_i, $i = \overline{1, m}$, such that the set A_1 is embedded in the half-plane $\Pi^\gamma(h_i)$, $i \le n_1$. In our case, we get $n_1 = 7$. We denote the point $\hat{e}_8^*$ by $\hat{e}_2(\vartheta)$.

Next, starting from the point $\hat{e}_2(\vartheta)$, we move on the part of the polyline γ, which lies inside the set A_1. In the set $h_j, j = \overline{n_1 + 1, m}$, we find a normal with the maximal number n_2, such that the intersection $\Phi_2^{n_2} = \cap_{i=n_1+1}^{n_2} \Pi^\gamma(h_i)$ of half-planes $\Pi^\gamma(h_i)$ is unbounded set. We get $n_2 = 11$ and denote the point $\check{e}_2(\vartheta)$.

Write $A_2 = A_1 \cap \Phi_2^{n_2}$. The boundary of the set A_2 consists of a polyline between the points $\hat{e}_2(\vartheta)$ and $\check{e}_2(\vartheta)$ (when one moves counterclockwise from the point $\hat{e}_2(\vartheta)$) and a part of the boundary of the set A_1 between the points $\check{e}_2(\vartheta)$ and $\hat{e}_2(\vartheta)$ (when one moves counterclockwise from the point $\check{e}_2(\vartheta)$).

In our case, the vector h_{n_2+1} is a normal to the set $-P$. We define the vertex $\hat{e}_3(\vartheta)$ of the polyline γ.

A part of the polyline γ, starting from the vertex $\hat{e}_3(\vartheta)$, lies inside the set A_2. We get that $n_3 = m = 14$, and we define the set A_3. The boundary of the set A_3 consists of a two-edge polyline between the points $\hat{e}_3(\vartheta)$ and $\check{e}_3(\vartheta)$ (when one moves counterclockwise from the point $\hat{e}_3(\vartheta)$) and a two-edge polyline on the boundary of the set A_2 between the points $\check{e}_3(\vartheta)$ and $\hat{e}_3(\vartheta)$ (when one moves counterclockwise from the point $\check{e}_3(\vartheta)$).

7.5.2 Construction of a Semipermeable Spiral Surface of Type ± for a Single Left Spiral

For a single left spiral γ, let us define a spiral system $\mathscr{A}^l(\gamma, P) = \{A_1 \supset A_2 \supset \cdots \supset A_n\}$. Consider the non-degenerated case $n \ge 2$. Let $a_\gamma(\vartheta)$ and $b_\gamma(\vartheta)$ be the start and end points of γ [twisting from the start point to the end one is *counterclockwise* (for the left spiral)].

It is known (Pshenichnyy and Sagaydak 1971) that for the problem of A_i-attainability ($i = \overline{1,n}$) the maximal u-stable bridge W_i is defined by

$$W_i := \{ (t,x) : t \in [0,\vartheta], x \in T_{\vartheta-t}(A_i;P,Q) \},$$

where

$$T_\tau(A;P,Q) := T_\tau(A) = \bigcap_{q\in Q}(A - \tau(P+q)), \quad \tau = \vartheta - t.$$

The lateral surfaces of the bridges W_i, $i = \overline{1,n}$, are semipermeable tubes. Using them, we construct a semipermeable spiral surface σ of type $\pm$ such that $\sigma(\vartheta) = \gamma$.

We define the length $\tau_* = \tau_*(\gamma)$ of a time interval for construction of the semipermeable surface σ by the condition $\operatorname{int} T_\tau(A_n) \neq \varnothing$, $\tau \in (0,\tau_*]$. Since A_n is a non-degenerated polygon (not a point, and not a segment), then such a value $\tau_* > 0$ exists. Write $t_* = t_*(\gamma) = \vartheta - \tau_*$.

Let Γ_i be the lateral surface of the bridge W_i on $[t_*,\vartheta]$, and let

$$\Gamma_i^\#(t) := \Gamma_i(t) \cap \Gamma_{i-1}(t), \quad i = \overline{2,n}.$$

The surfaces Γ_i, $i = \overline{1,n}$, are semipermeable tubes of type $\pm$.

By Lemma 7, for $i = \overline{2,n}$, $t \in [t_*,\vartheta]$, $\tau = \vartheta - t$, we have the representation

$$T_\tau(A_i) = T_\tau(A_{i-1}) \bigcap \left(\bigcap \{T_\tau(\Pi) : \Pi \in \mathscr{R}_P(A_i) \setminus \mathscr{R}_P(A_{i-1})\} \right).$$

Here, the set $\mathscr{R}_P(A_i)$ of half-planes is defined by (7.7). In view of definition of A_i on the basis of A_{i-1} and the value n_i, we have

$$\Phi_i^{n_i} = \bigcap \{\Pi : \Pi \in \mathscr{R}_P(A_i) \setminus \mathscr{R}_P(A_{i-1})\}.$$

The set $\Phi_i^{n_i}$ is given by (7.9) and unbounded. Due to property (4) of the operator T_τ, the image $T_\tau(\Pi)$ of the half-plane Π is its shift by the vector $-\tau(p_* + q_*)$. Therefore, the intersection

$$\bigcap \{T_\tau(\Pi) : \Pi \in \mathscr{R}_P(A_i) \setminus \mathscr{R}_P(A_{i-1})\}$$

of the half-planes is also unbounded. So, since the set A_i is non-degenerated, we deduce that the set $\Gamma_i^\#(t)$ is polyline (either closed or nonclosed) and the set-valued function $t \to \Gamma_i^\#(t)$ is continuous.

Using symbols $\hat{e}_i(t)$, $\check{e}_i(t)$, we denote the vertices of the polyline $\Gamma_i^\#(t)$ (if it is nonclosed) in such a way that the set $W_i(t)$ is on the left hand when moving along $\Gamma_i^\#(t)$ from $\check{e}_i(t)$ to $\hat{e}_i(t)$, $i = \overline{2,n}$. The functions $\hat{e}_i(\cdot)$ and $\check{e}_i(\cdot)$ are continuous. The introduced notations are explained in Fig. 7.4 for $t = \vartheta$.

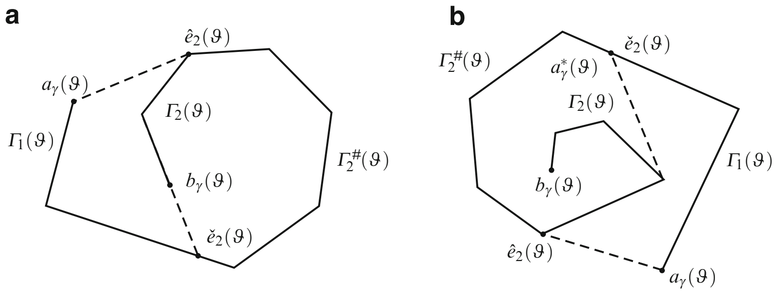

Fig. 7.4 Construction of a semipermeable spiral surface: the basis (**a**) and the inductive step (**b**)

Choose some disjoint continuous trajectories $a_\gamma(t)$, $b_\gamma(t)$, $t \in [t_*, \vartheta]$, with the end points $a_\gamma(\vartheta)$, $b_\gamma(\vartheta)$, such that the trajectory $a_\gamma(\cdot)$ lies on the tube Γ_1, and the trajectory $b_\gamma(\cdot)$ lies on Γ_n. We do not impose any geometric restrictions on location of the curves $a_\gamma(\cdot)$ and $b_\gamma(\cdot)$ with respect to the surface $\Gamma_i^{\#}$.

Now we turn to the construction (by math induction on n) of a semipermeable spiral surface σ of type $\pm$ (with the edges $a_\gamma(\cdot)$, $b_\gamma(\cdot)$).

(1) The basis: let $n = 2$ (Fig. 7.4a). We have two semipermeable tubes Γ_1 and Γ_2.

(1a) First, suppose that $W_1(t) \neq W_2(t)$ for all $t \in [t_*, \vartheta)$. In this case, the disjoint trajectories $\hat{e}_2(\cdot)$ and $\check{e}_2(\cdot)$ (the edges of the surface $\Gamma_2^{\#}$) are given on the whole segment $[t_*, \vartheta]$.

Define disjoint trajectories $\hat{\xi}_i(\cdot)$, $\check{\xi}_i(\cdot)$ that lie on the tube Γ_i, $i = 1, 2$. For $t \in [t_*, \vartheta]$, let

$$\begin{aligned}
\check{\xi}_1(t) &= \begin{cases} \hat{e}_2(t) & \text{if } b_\gamma(t) \notin \Gamma_2^{\#}(t) \\ b_\gamma(t) & \text{otherwise} \end{cases}, \qquad \hat{\xi}_1(t) = a_\gamma(t), \\
\hat{\xi}_2(t) &= \begin{cases} \check{e}_2(t) & \text{if } a_\gamma(t) \notin \Gamma_2^{\#}(t) \\ a_\gamma(t) & \text{otherwise} \end{cases}, \qquad \check{\xi}_2(t) = b_\gamma(t).
\end{aligned} \tag{7.10}$$

Since the trajectories $a_\gamma(\cdot)$, $b_\gamma(\cdot)$, $\hat{e}_2(\cdot)$, and $\check{e}_2(\cdot)$ are continuous, and $\hat{e}_2(\cdot)$, $\check{e}_2(\cdot)$ are the edges of the surface $\Gamma_2^{\#}$, we conclude that the trajectories $\hat{\xi}_i(\cdot)$ and $\check{\xi}_i(\cdot)$ are also continuous.

Write

$$\sigma_i(\cdot) = S^+(\cdot\,; \Gamma_i, [t_*, \vartheta], \hat{\xi}_i(\cdot), \check{\xi}_i(\cdot)), \quad i = 1, 2; \quad \sigma = \sigma_1 \cup \sigma_2.$$

The tubes Γ_1 and Γ_2 are semipermeable. So in view of Lemma 2 we obtain that the surfaces $\sigma_1 \subset \Gamma_1$ and $\sigma_2 \subset \Gamma_2$ are semipermeable. Note that $\sigma_1 \cap \sigma_2$ is a semipermeable surface with the edges $\hat{\xi}_2(\cdot)$ and $\check{\xi}_1(\cdot)$, and $\sigma = \sigma_1 \cup \sigma_2$ is a surface with the edges $\hat{\xi}_1(\cdot)$, $\check{\xi}_2(\cdot)$.

Since $W_2(t) \subset W_1(t)$ and Γ_1, Γ_2 are semipermeable tubes of type $\pm$, the sides $(+)$ and $(-)$ of the surfaces σ_1 and σ_2 define the sides $(+)$ and $(-)$ of the surface σ. By Lemma 3, the surface σ is semipermeable surface of type $\pm$ with definite sides $(+)$ and $(-)$.

(1b) Consider the remaining case: there exists $\bar{t} \in [t_*, \vartheta)$ such that $W_1(\bar{t}) = W_2(\bar{t})$. In this case, $W_1(t) = W_2(t)$ for $t \in [t_*, \bar{t}\,]$.

For $t \in (\bar{t}, \vartheta)$, two different values $\hat{e}_2(t)$ and $\check{e}_2(t)$ are given, which are the ends of the polyline $\Gamma^{\#}(t)$. Therefore we define the values $\hat{\xi}_i(t)$ and $\check{\xi}_i(t)$ by (7.10), $i = 1, 2$.

For $t \in [t_*, \bar{t}\,]$, we have $\Gamma_1(t) = \Gamma_2(t)$. Therefore we define $\hat{\xi}_1(t) = \hat{\xi}_2(t) = a_\gamma(t)$ and $\check{\xi}_1(t) = \check{\xi}_2(t) = b_\gamma(t)$.

We next prove that the trajectories $\hat{\xi}_i(\cdot)$ and $\check{\xi}_i(\cdot)$ are continuous on $[t_*, \vartheta]$, $i = 1, 2$. In virtue of definition of the functions $\hat{\xi}_i(\cdot)$ and $\check{\xi}_i(\cdot)$, we are reduced to proving right continuity of the functions $\check{\xi}_1(\cdot)$ and $\hat{\xi}_2(\cdot)$ at $t = \bar{t}$.

Define

$$a_* = \lim_{t \to \bar{t}+0} \hat{e}_2(t) = \lim_{t \to \bar{t}+0} \check{e}_2(t).$$

The second equality is not trivial and based on Lemma 9. The proof is omitted. We have $a_* \in \Gamma_2(\bar{t}) = \Gamma_1(\bar{t})$.

Let us begin with the function $\check{\xi}_1(\cdot)$. Assume that there exists a sequence $\{t_k\}$ such that $t_k \to \bar{t} + 0$ and $\check{\xi}_1(t_k) = \hat{e}_2(t_k) \not\to b_\gamma(\bar{t})$. Then $b_\gamma(t_k) \in \Gamma_2(t_k) \setminus \Gamma_2^{\#}(t_k)$. But the polyline $\Gamma_2(t_k) \setminus \Gamma_2^{\#}(t_k)$ is collapsed to the point a_*, i.e. $b_\gamma(t_k) \to a_*$. Therefore, $b_\gamma(\bar{t}) \neq a_*$ and the point $b_\gamma(\bar{t})$ does not belong to a neighbourhood G_* of a_*. The points $\hat{e}_2(t_k)$ and $\check{e}_2(t_k)$ are required to get into the neighbourhood G_* for sufficiently large k. So given continuity of the trajectory $b_\gamma(\cdot)$, we obtain $b_\gamma(t_k) \in \Gamma_2^{\#}(t_k)$, i.e. $\hat{\xi}_1(t_k) = b_\gamma(t_k)$, that contradicts our assumption. Right continuity of $\check{\xi}_1(\cdot)$ at $t = \bar{t}$ is proved.

Now consider the function $\hat{\xi}_2(\cdot)$ and assume that there exists a sequence $\{t_k\}$ such that $t_k \to \bar{t} + 0$ and $\hat{\xi}_2(t_k) = \check{e}_2(t_k) \not\to a_\gamma(\bar{t})$. Then $a_\gamma(t_k) \in \Gamma_1(t_k) \setminus \Gamma_2(t_k)$. Since the set $\Gamma_1(t_k) \setminus \Gamma_2(t_k)$ is collapsed to the point a_*, we have $a_\gamma(t_k) \to a_*$. At the same time, we observe $a_\gamma(t_k) \to a_\gamma(\bar{t})$ and $\check{e}_2(t_k) \to a_*$. Thus $a_\gamma(\bar{t}) = a_*$ and $\check{e}_2(t_k) \to a_\gamma(\bar{t})$, that contradicts our assumption. We get that $\hat{\xi}_2(\cdot)$ is also right continuous at $t = \bar{t}$.

Further arguments in case (1b) are the same as for case (1a).

(2) The inductive step: assume that we have an algorithm to construct a semipermeable surface for $n = k$, $k \geq 2$; let us constructs it for $n = k + 1$.

Write the sequence $\Gamma_1, \Gamma_2, \ldots \Gamma_k, \Gamma_{k+1}$ of semipermeable tubes on the time interval $[t_*, \vartheta]$. Consider two possible cases like in the proof of the basis.

(2a) First assume that $W_1(t) \neq W_2(t)$ for all $t \in [t_*, \vartheta)$. In this case, the disjoint trajectories $\hat{e}_2(\cdot)$ and $\check{e}_2(\cdot)$ (the edges of the surface $\Gamma_2^{\#}$) are given on the whole interval $[t_*, \vartheta]$. For $t \in [t_*, \vartheta]$, we write

$$a_\gamma^*(t) = \begin{cases} \check{e}_2(t) & \text{if } a_\gamma(t) \notin \Gamma_2^\#(t) \\ a_\gamma(t) & \text{otherwise.} \end{cases} \tag{7.11}$$

Figure 7.4b illustrates the notations for $t = \vartheta$. Note that the trajectory $a_\gamma^*(\cdot)$ is defined in the same way as the function $\hat{\check{\xi}}_2(\cdot)$ in item (1b), and therefore it is continuous as well.

Applying the induction assumption for the sequence of k tubes $\Gamma_2, \ldots, \Gamma_k$, and Γ_{k+1}, we construct the semipermeable surface σ_* with edge trajectories $a_\gamma^*(\cdot)$ and $b_\gamma(\cdot)$.

The trajectories $a_\gamma(\cdot)$ and $a_\gamma^*(\cdot)$ lie on the tube Γ_1 and are continuous. Write

$$\sigma_1(\cdot) = S^+(\,\cdot\,; \Gamma_1, [t_*, \vartheta], a_\gamma(\cdot), \hat{e}_2(\cdot)), \quad \sigma = \sigma_1 \cup \sigma_*.$$

Using Lemma 3 for the surfaces σ_1 and σ_*, we find that the surface σ is semipermeable.

(2b) Assume that there exists $\bar{t} \in [t_*, \vartheta)$ such that $W_1(\bar{t}) = W_2(\bar{t})$. Then define the value $a_\gamma^*(t)$ by formula (7.11) for $t \in (\bar{t}, \vartheta]$, and let $a_\gamma^*(t) = a_\gamma(t)$ for $t \in [t_*, \bar{t}]$.

We construct the semipermeable surface on $(\bar{t}, \vartheta]$ in the same way as in case (2a). On the interval $[t_*, \bar{t}]$, the semipermeable surface can be defined by the induction assumption applied for the tubes $\Gamma_2, \ldots, \Gamma_{k+1}$ since $\Gamma_1(t) = \Gamma_2(t)$, $t \in [t_*, \bar{t}]$. On the whole interval $[t_*, \vartheta]$, we get a semipermeable surface.

7.5.3 Construction of Semipermeable Spiral Surfaces of Types ± and ∓ for Other Original Spirals

(a) When constructing a semipermeable spiral surface of type $\pm$ for a single right spiral, we define the right spiral system $\mathscr{A}^r(\gamma, P)$ by ordering all the normals *clockwise* and moving along boundaries of the sets in the negative direction (the set is on the right hand). We compose the spiral surface in the same way as for a single left spiral replacing the operator $S^+(\cdot)$ by $S^-(\cdot)$.

(b) When constructing a semipermeable spiral surface of type $\mp$ for a single left (right) spiral, we define the spiral system $\mathscr{A}^l(\gamma, Q)$ (correspondingly, $\mathscr{A}^r(\gamma, Q)$). We compose the spiral surface of type $\mp$ in the same way as in the construction of a semipermeable spiral surface of type $\pm$ replacing the operator $T_\tau(\cdot; P, Q)$ by $T_\tau(\cdot; Q, P)$, and the set $\mathscr{R}_P(A)$ by $\mathscr{R}_Q(A)$.

(c) Let us describe a construction of semipermeable spiral surfaces of types $\pm$ for a double spiral γ. By definition, we have the representation $\gamma = \gamma_0{}^l \cup \gamma_0 \cup \gamma_0{}^r$, where $\gamma_0{}^l$, γ_0, $\gamma_0{}^r$ are non-degenerated polylines,

$$\gamma_0 = \gamma \cap \partial A_0, \quad A_0 = \mathrm{co}\,\gamma, \quad \gamma_0^{\,l} \cap \gamma_0^{\,r} = \varnothing,$$

$\gamma_0 \cup \gamma_0^{\,l} = \gamma^{\,l}$ is a left spiral, $\gamma_0 \cup \gamma_0^{\,r} = \gamma^{\,r}$ is a right spiral (Fig. 7.1c).

Let $a_\gamma^l(\vartheta) = \gamma_0 \cap \gamma_0^{\,r}$ and $a_\gamma^r(\vartheta) = \gamma_0 \cap \gamma_0^{\,l}$ be initials vertices of the spirals $\gamma^{\,l}$ and $\gamma^{\,r}$, and let $b_\gamma^l(\vartheta)$ and $b_\gamma^r(\vartheta)$ be its end vertices. Note that $b_\gamma^l(\vartheta) \in \gamma_0^{\,l}$ and $b_\gamma^r(\vartheta) \in \gamma_0^{\,r}$ are the end vertices of the spiral γ (the both are "inner").

Let trajectories $a_\gamma^l(\cdot)$ and $a_\gamma^r(\cdot)$ be given. For the left spiral $\gamma^{\,l}$ and right spiral $\gamma^{\,r}$, we define the corresponding left and right spiral systems:

$$\mathscr{A}^l(\gamma^{\,l}) = \{A_1^l \supset A_2^l \supset \cdots \supset A_{n_l}^l\}, \qquad \mathscr{A}^r(\gamma^{\,r}) = \{A_1^r \supset A_2^r \supset \cdots \supset A_{n_r}^r\}.$$

We have $A_1^l = A_1^r = A_0$.

Choose an instant $t_*^{l,r} < \vartheta$ such that $\mathrm{int}\, T_\tau(A_{n_l}^l) \neq \varnothing$ and $\mathrm{int}\, T_\tau(A_{n_r}^r) \neq \varnothing$ for $\tau \in (0, \vartheta - t_*^{l,r}]$.

Let W_0, W_2^l, W_2^r be the maximal u-stable bridges in the problems of A_0-, A_2^l-, A_2^r-attainability, correspondingly; and let Γ_0, Γ_2^l, Γ_2^r denote the lateral surfaces of the bridges W_0, W_2^l, W_2^r.

We define a trajectory $a_\gamma^l(\cdot)$ [correspondingly, $a_\gamma^r(\cdot)$] as the edge (along the time axis) of the set $\overline{\Gamma_0 \setminus \Gamma_2^r}$ [correspondingly, $\overline{\Gamma_0 \setminus \Gamma_2^l}$], moving backward from $a_\gamma^l(\vartheta)$ [correspondingly, $a_\gamma^r(\vartheta)$].

Find an instant $\tilde{t} \in [t_*^{l,r}, \vartheta)$ such that the trajectories $a_\gamma^l(\cdot)$ and $a_\gamma^r(\cdot)$ are not intersect on $[\tilde{t}, \vartheta]$. Write

$$\sigma_0(\cdot) = S^+(\,\cdot\,; \Gamma_0, [\tilde{t}, \vartheta], a_\gamma^l(\cdot), a_\gamma^r(\cdot)).$$

On $[\tilde{t}, \vartheta]$, assume that trajectories $b_\gamma^l(\cdot)$ and $b_\gamma^r(\cdot)$, that are the "inner" edges of the semipermeable surface to be constructed, are given, and define the semipermeable surfaces σ^l and σ^r of type $\pm$ with the edges $a_\gamma^l(\cdot)$, $b_\gamma^l(\cdot)$ and $a_\gamma^r(\cdot)$, $b_\gamma^r(\cdot)$, correspondingly. Let the value $t^* = t^*(\gamma) \in [\tilde{t}, \vartheta)$ be defined by the condition of non-intersection of sets $\sigma^l(t)$ and $\sigma^r(t)$ outside the set $\sigma_0(t)$ for $t \in [t^*, \vartheta]$.

The surface $\sigma = \sigma^l \cup \sigma^r$ is the desired semipermeable surface.

7.6 Sewing a Semipermeable Tube from Semipermeable Spiral Surfaces

Consider a polygonal boundary $\gamma = \partial M$ of the nonconvex terminal set. We represent the closed polyline γ in the form of an ordered union of even number of spirals γ_i, $i = \overline{1, 2k}$, such that the last edge of a spiral coincides with the first edge of the next spiral (common edges) while the corresponding normals to the common edges are in opposite directions. Several variants of the set M is given in Fig. 7.5. Common edges of the spirals are shown by bold lines.

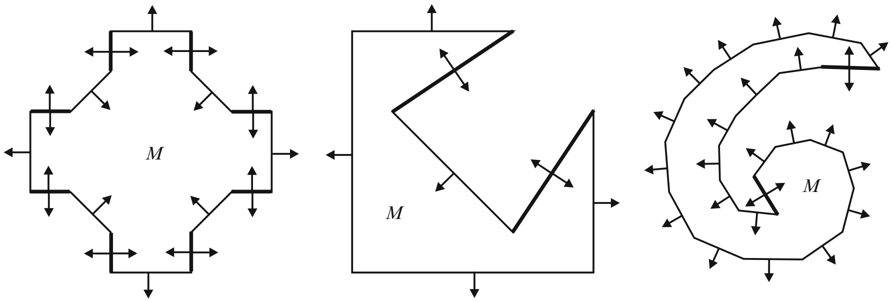

Fig. 7.5 Partition of the boundary of the terminal set M into convex and concave polyline arcs intersecting by common edges

To be definite, assume that the normals of the spiral γ_i are outward the set M for an odd i (*a convex arc* on the boundary of the set). In this case, we construct a semipermeable spiral surface of type $\pm$. For an even i, we assume that the normals are inward the set M (*a concave arc*). For such a spiral, we construct a semipermeable spiral surface of type $\mp$.

Let each single spiral γ_i be assigned to the value $t^\flat(\gamma_i) = t_*(\gamma_i)$, where $t_*(\gamma_i)$ is introduced in Sect. 7.5.2. In the case of a double spiral, we write $t^\flat(\gamma_i) = t^*(\gamma_i)$, where $t^*(\gamma_i)$ is introduced at the end of Sect. 7.5.3. On $[t^\flat(\gamma_i), \vartheta]$, we construct a semipermeable surface σ_i emanating backward in time from the spiral γ_i, i.e. $\sigma_i(\vartheta) = \gamma_i$.

Find

$$t_M \geq \max_{i \in \overline{1,2k}} t^\flat(\gamma_i),$$

such that the set $\sigma = \sigma_1 \cup \cdots \cup \sigma_{2k}$ is a tube on $[t_M, \vartheta]$. The value t_M is defined by the following conditions: (1) the common edges do not degenerate; (2) sections $\sigma_i(t)$, $t \in [t_M, \vartheta]$, can intersect with adjacent surfaces only by the common edges. Condition (1) means that the common edge of any two adjacent convex and concave arcs of the original polyline γ changes its length in a section for $t \in [t_M, \vartheta]$, when the corresponding two spiral surfaces go backwards in time, but the common edge does not collapse to a point.

Using Lemmas 3 and 4, we deduce that the tube σ is semipermeable on $[t_M, \vartheta]$, and $\sigma(\vartheta) = \gamma$. Consequently, in virtue of Lemma 1, the tube σ coincides with the lateral surface of the maximal u-stable bridge W_0 on $[t_M, \vartheta]$, i.e. the tube σ gives the solution of the differential game on the segment.

7.7 Example

We illustrate our theoretical results by the following example. Let a nonconvex set M and geometrical constraints P and Q for controls of the players be given (Fig. 7.6). Write $\vartheta = 1$.

The boundary of M is divided into the four convex curves: one left spiral γ_1 and three ordinary convex curves γ_2, γ_3, γ_4. In Fig. 7.7, a schematic images of the curves are shown by dotted lines, where each dotted line is shifted from the corresponding line on the boundary of M. The arcs γ_1 and γ_3 are convex arcs on the boundary of M, and the arcs γ_2 and γ_4 are concave. Define $M' = \overline{\mathbb{R}^2 \setminus M}$.

The result of the construction of the section $W_0(0)$ of the maximal u-stable bridge W_0 at $t = 0$ is shown in Fig. 7.8 by a solid line 2, the boundary of the terminal set $W_0(\vartheta) = M$ is given by a polyline 1.

To make a comparison, we construct the sets $T_\vartheta(M; P, Q)$ (dashed line 3) and $T_\vartheta(M'; Q, P)$ (dotted line 4). One can see that the set $W_0(0)$ is smaller than the first one, but it is larger than the closure of the complement of the second one. The same property is valid for any $t \in [0, 1)$ if we compute the sets $W_0(t)$, $T_{\vartheta-t}(M; P, Q)$, and $T_{\vartheta-t}(M'; Q, P)$.

Thus we have an example of a game problem of guidance with simple motions, fixed terminal instant $\vartheta = 1$, and nonconvex terminal set M, for which the solvability set is constructed on $[0, 1]$ by the method described above. To construct the t-section $W_0(t)$, the method does not require an additional partition of the interval $[t, \vartheta]$. The set $W_0(t)$ does not coincide with the set of program absorption by the first player for the set M. The closure of its complement $\overline{\mathbb{R}^2 \setminus W_0(t)}$ does not coincide with the set of program absorption by the second player for the set $\overline{\mathbb{R}^2 \setminus M}$.

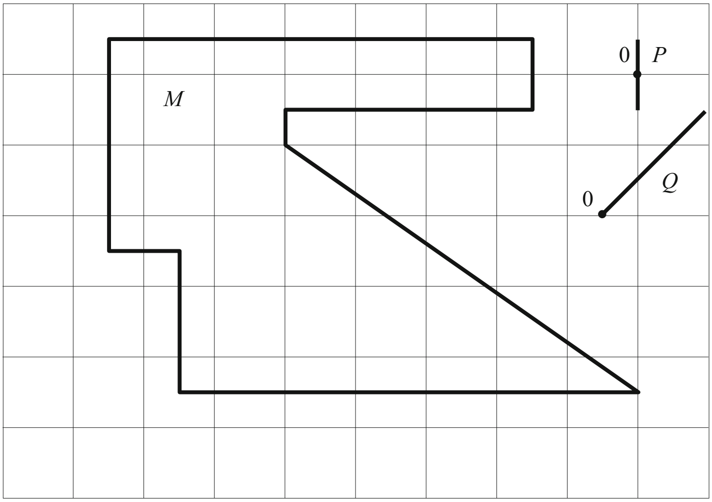

Fig. 7.6 Example: the terminal set M, the constraints P and Q for controls of the players

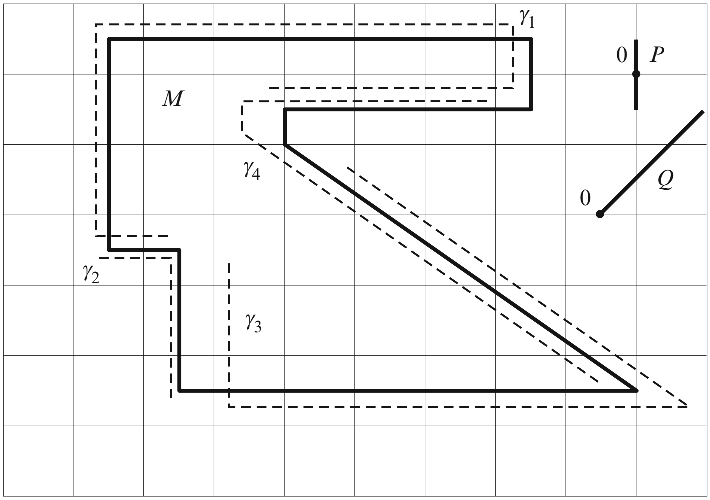

Fig. 7.7 Example: the partition of the boundary of M into convex spirals γ_1, γ_2, γ_3, and γ_4

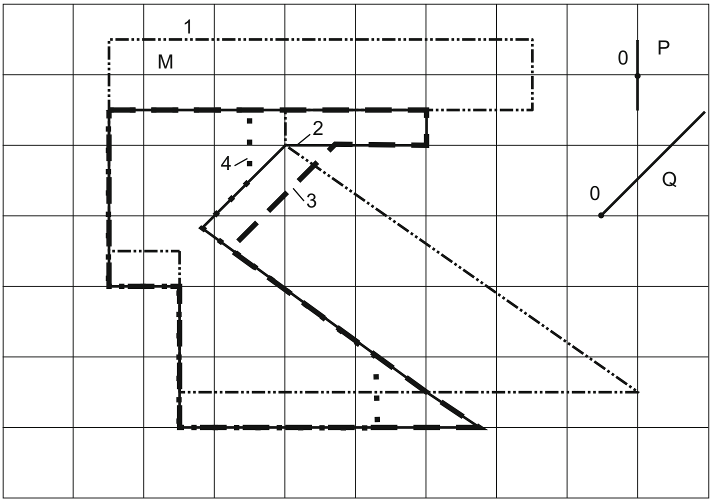

Fig. 7.8 Example: *1* denotes the boundary of $M = W_0(\vartheta)$; *2* denotes the boundary of $W_0(0)$; *3* denotes the boundary of $T_\vartheta(M; P, Q)$; *4* denotes the boundary of $T_\vartheta(M'; Q, P)$, $M' = \overline{\mathbb{R}^2 \setminus M}$; $\vartheta = 1$

7.8 Conclusion

In the article, a zero-sum differential game with simple motions in the plane and a fixed terminal time is considered. For a polygonal terminal set and convex polygonal constraints on the players' controls, it is shown that there exists a time interval such that it is adjacent to the terminal instant and the solvability set (maximal stable bridge) on this interval can be constructed by open-loop controls without any additional partitions and passages to the limits afterwards. In this paper, there is no any lower estimate of the length of the time interval.

Acknowledgements We are grateful to the unknown reviewer for helpful remarks.

The research was carried out under financial support of the Russian Foundation for Basic Research (project no. 15-01-07909).

Appendix

Proof (of Lemma 5). We define

$$r_0 := \operatorname{dist}(A, \partial \Pi_0), \quad \Pi_1 := \Pi_0 - r_0 \nu_0 = \Pi(\alpha_0 - r_0, \nu_0).$$

Then $A \subset \Pi_1 \subset \Pi_0$. Using (7.6), we get

$$T_\tau(\Pi_1) = T_\tau(\Pi_0) - r_0 \nu_0.$$

Since $A \subset \Pi_1$, we have $T_\tau(A) \subset T_\tau(\Pi_1)$. Consequently,

$$\operatorname{dist}(T_\tau(A), \partial T_\tau(\Pi_0)) = \operatorname{dist}(T_\tau(A), \partial T_\tau(\Pi_1)) + r_0 \geq r_0.$$

Proof (of Lemma 6). Since A is a convex polygon, we have

$$A = \bigcap \{\Pi(\rho(\eta, A), \eta) : \ \eta \in \mathscr{N}(A)\}. \tag{7.12}$$

We add extra half-planes supporting A with the outer normals from $\mathscr{N}(-P)$ to the intersection (7.12):

$$A = \bigcap \{\Pi(\rho(\eta, A), \eta) : \ \eta \in \mathscr{N}(A) \cup \mathscr{N}(-P)\}.$$

Then by (7.5), we observe

$$T_\tau(A) \subset \bigcap \{T_\tau(\Pi(\rho(\eta, A), \eta)) : \ \eta \in \mathscr{N}(A) \cup \mathscr{N}(-P)\} =: Y.$$

We prove the opposite inclusion by contradiction. Assume that we can choose $y \in Y \setminus T_\tau(A)$. Since $y \notin T_\tau(A)$, there exists $\tilde{q} \in Q$ such that $(y + \tau(P + \tilde{q})) \cap A = \varnothing$, i.e. we have $(z + \tau P) \cap A = \varnothing$ for $z = y + \tau \tilde{q}$.

So, there exists $\eta_* \in \mathscr{N}(A) \cup \mathscr{N}(-P)$ such that

$$\Pi(\rho(\eta_*, A), \eta_*) \cap (z + \tau P) = \varnothing. \tag{7.13}$$

Property (4) of the operator T_τ implies the relation

$$y \in T_\tau(\Pi(\rho(\eta_*, A), \eta_*)) = \Pi(\rho(\eta_*, A), \eta_*) - \tau(p_* + q_*),$$

where

$$p_* \in \operatorname{Arg}\min_{p \in P} \langle p, \eta_* \rangle, \quad q_* \in \operatorname{Arg}\max_{q \in Q} \langle q, \eta_* \rangle.$$

Thus $y + \tau(p_* + q_*) \in \Pi(\rho(\eta_*, A), \eta_*)$, i.e. $\langle y + \tau(p_* + q_*), \eta_* \rangle \leq \rho(\eta_*, A)$.

By definition of q_*, for any $q \in Q$ the inequality $\langle q, \eta_* \rangle \leq \langle q_*, \eta_* \rangle$ holds. Consequently, $\langle y + \tau(p_* + \tilde{q}), \eta_* \rangle \leq \rho(\eta_*, A)$, which implies $y + \tau\tilde{q} + \tau p_* \in \Pi(\rho(\eta_*, A), \eta_*)$. Since $p_* \in P$, we observe

$$\Pi(\rho(\eta_*, A), \eta_*) \cap (y + \tau\tilde{q} + \tau P) \neq \varnothing.$$

Using the definition of the point z, we get a contradiction to (7.13).

Proof (of Lemma 7). In virtue of Lemma 6, we have

$$T_\tau(A) = \bigcap \{T_\tau(\Pi) : \Pi \in \mathscr{R}_P(A)\}, \quad T_\tau(B) = \bigcap \{T_\tau(\Pi) : \Pi \in \mathscr{R}_P(B)\},$$

where the sets $\mathscr{R}_P(A)$, $\mathscr{R}_P(B)$ of half-planes are defined by (7.7). Hence, using the inclusion of $B \subset A$, we obtain the representation (7.8).

Proof (of Lemma 8). Let E denote the set of all $\tau_1 \in (0, \tau_0)$ such that for any $\tau \in [\tau_1, \tau_0]$ the set $T_\tau(A) \cap \partial T_\tau(\Pi)$ is a non-degenerated segment (an edge of the polygon $T_\tau(A)$). Property (4) of the operator T_τ and Lemma 6 imply that the set $E \subset (0, \tau_0)$ is nonempty and open.

Let $\bar{\tau} = \inf E$. The assertion of this lemma is equivalent to the equality $\bar{\tau} = 0$. Assume $\bar{\tau} > 0$. Since the set E is open, we observe that the set $T_{\bar{\tau}}(A) \cap \partial T_{\bar{\tau}}(\Pi) =: \{\bar{a}\}$ is a singleton.

Write $\Pi = \Pi(\alpha, \nu)$. Define a straight line L_ν which is parallel to the vector ν and passes through the point $\bar{a}$.

In virtue of Lemma 6, the edges of the polygon $T_{\bar{\tau}}(A)$ that are adjacent to the vertex $\bar{a}$ lie on the boundary of the half-planes $T_{\bar{\tau}}(\Pi_1)$ and $T_{\bar{\tau}}(\Pi_2)$ for some $\Pi_1, \Pi_2 \in \mathscr{R}_P(A)$. Let $a_{1,2}(\tau)$ denote the orthogonal projection onto the line L_ν of the point of intersection of two lines $\partial T_\tau(\Pi_1)$ and $\partial T_\tau(\Pi_2)$. Let $a(\tau)$ denote the point of projection of $\partial T_\tau(\Pi)$ onto L_ν. We observe that

$$a(\bar{\tau}) = a_{1,2}(\bar{\tau}) = \bar{a}, \quad \langle a(\tau), \nu \rangle < \langle a_{1,2}(\tau), \nu \rangle, \quad \tau > \bar{\tau}.$$

Property (4) of the operator T_τ implies that the velocity of the points $a(\tau)$ and $a_{1,2}(\tau)$ along the line L_ν is constant with respect to τ. Consequently,

$$\langle a(\tau), \nu \rangle > \langle a_{1,2}(\tau), \nu \rangle, \quad \tau < \bar{\tau}$$

therefore

$$T_\tau(A) \cap \partial T_\tau(\Pi) = \varnothing, \quad \tau \in (0, \bar{\tau}).$$

This relation contradicts to Lemma 5 since

$$\text{dist}\,(T_\tau(A), \partial T_\tau(\Pi)) > 0, \quad \text{dist}\,(T_{\bar{\tau}}(A), \partial T_{\bar{\tau}}(\Pi)) = 0, \quad \tau \in (0, \bar{\tau}).$$

Proof (of Lemma 9). We prove the assertion by contradiction. For brevity, we denote

$$\Gamma^{\#}(\tau) := \partial T_\tau(A) \cap \partial T_\tau(B), \quad \tau \geq 0.$$

Assume that the set $\Gamma^{\#}(\tau)$ is nonempty and disconnected.

Lemma 7 [representation (7.8)] implies that a connected component of $\Gamma^{\#}(\tau)$ cannot "split", and disconnectedness can appear only with a new intersection. Since $\Gamma^{\#}(0) = \partial A \cap \partial B$ is a nondegenerate nonclosed polyline, there exists the first instant $\tau_0 > 0$ of disconnectedness. Consequently, we can find a point $\xi_0 \in \Gamma^{\#}(\tau_0)$, a neighbourhood $G_0 \ni \xi_0$, and a value $\delta_0 \in (0, \tau_0)$ such that

$$G_0 \cap \Gamma^{\#}(\tau_0 - \delta) = \varnothing, \quad \delta \in (0, \delta_0). \tag{7.14}$$

Let η_0 denote the unit outer normal to those edge of the polygon $T_{\tau_0}(A)$ that contains the point ξ_0. Write

$$\Pi_{\tau_0} = \Pi(\rho(\eta_0, T_{\tau_0}(A)), \eta_0), \quad \Pi_\tau = \Pi_{\tau_0} + (\tau_0 - \tau)(p_* + q_*), \quad \tau \in [0, \tau_0],$$

where

$$p_* \in \text{Arg} \min_{p \in P} \langle p, \nu \rangle, \quad q_* \in \text{Arg} \max_{q \in Q} \langle q, \nu \rangle.$$

Property (4) of the operator T_τ implies

$$\Pi_\tau = T_\tau(\Pi_0), \quad \tau \in [0, \tau_0].$$

The half-plane Π_{τ_0} is supporting to the sets $T_{\tau_0}(A)$ and $T_{\tau_0}(B)$.

Since $T_{\tau_0}(A) \cap \partial \Pi_{\tau_0}$ is an edge of the polygon $T_{\tau_0}(A)$, Lemma 8 implies that the set $T_\tau(A) \cap \partial \Pi_\tau$ is an edge of the polygon $T_\tau(A)$ for $\tau \in [0, \tau_0)$.

We deduce that the set $\partial \Pi_{\tau_0} \cap T_{\tau_0}(B)$ is not an edge of the polygon $T_{\tau_0}(B)$, otherwise we get a contradiction with (7.14) by Lemma 8. As a result, we have $\partial \Pi_{\tau_0} \cap T_{\tau_0}(B) = \{\xi_0\}$. Using Lemma 8 and relation (7.14), we observe that $\partial \Pi_\tau \cap T_\tau(B) = \varnothing$. Therefore,

$$\text{dist}\,(T_\tau(B), \partial \Pi_\tau) = r > 0.$$

Applying Lemma 5 for the operator $T_\delta(\cdot)$, $\delta := \tau_0 - \tau$, we find

$$\text{dist}\,(T_\delta(T_\tau(B)), \partial T_\delta(\Pi_\tau)) \geq r.$$

We have

$$T_\delta(T_\tau(B)) = T_{\tau_0}(B), \quad T_\delta(\Pi_\tau) = \Pi_{\tau_0}.$$

Since the half-plane Π_{τ_0} is supporting to the set $T_{\tau_0}(B)$, we write

$$\text{dist}\,(T_\delta(T_\tau(B)), \partial T_\delta(\Pi_\tau)) = 0,$$

that contradicts to the condition $r > 0$.

References

Botkin ND, Hoffman K-H, Turova VL (2011) Stable numerical schemes for solving Hamilton–Jacobi–Bellman–Isaacs equations. SIAM J Sci Comput 33(2):992–1007

Cardaliaguet P, Quincampoix M, Saint-Pierre P (1999) Set-valued numerical analysis for optimal control and differential games. In: Bardi M, Raghavan TE, Parthasarathy T (eds) Annals of the international society of dynamic games, vol 4. Birkhauser, Boston, pp 177–247

Cristiani E, Falcone M (2006) A fast marching method for pursuit-evasion games. In: Communications to SIMAI Congress, vol 1. doi:10.1685/CSC06059. ISSN 1827-9015

Dvurechensky PE, Ivanov GE (2014) Algorithms for computing Minkowski operators and their application in differential games. Comput Math Math Phys 54(2):235–264

Isaacs R (1965) Differential games. Wiley, New York

Kamneva LV, Patsko VS (2013) Open-loop solvability operator in differential games with simple motions in the plane. In: Krivan V, Zaccour G (eds) Advances in dynamic games: theory, applications, and numerical methods. Annals of the international society of dynamic games, vol 13. Birkhäuser, Basel, pp 239–262

Krasovskii NN, Subbotin AI (1974) Positional differential games. Nauka, Moscow (in Russian)

Krasovskii NN, Subbotin AI (1988) Game-theoretical control problems. Springer, New York

Kumkov SS, Patsko VS, Shinar J (2005) On level set with "narrow throats" in linear differential games. Int Game Theory Rev 7(3):285–311

Pshenichnyy BN, Sagaydak MI (1971) Differential games with fixed time. J Cybernet 1:117–135

Sethian JA (1999) Level set methods and fast marching methods. Cambridge University Press, Cambridge

Subbotin AI (1995) Generalized solutions of first-order PDEs: the dynamical optimization perspective. Birkhäuser, Boston

Subbotin AI, Patsko VS (eds) (1984) Algorithms and programs for solving linear differential games. Institute of Mathematics and Mechanics, Sverdlovsk (in Russian)

Taras'ev AM, Ushakov VN, Khripunov AP (1987) On a computational algorithm for solving game control problem. J Appl Math Mech 51(2):216–222 (in Russian)

Chapter 8
Linear-Quadratic Gaussian Dynamic Games with a Control-Sharing Information Pattern

Meir Pachter

Abstract A "zero-sum" Linear-Quadratic Gaussian Dynamic Game (LQGDG) where the players have partial information is considered. Specifically, the players' initial state information and their measurements are private information, but each player is able to observe his antagonist's past inputs: the protagonists' past controls is shared information. Although this is a game with partial information, the control-sharing information pattern renders the game amenable to solution by the method of dynamic programming. The correct solution of LQGDGs with a control-sharing information pattern is obtained in closed-form.

Keywords Linear quadratic Gaussian dynamic game • Partial information

MSC Codes: 91A25, 93C41, 49N70

8.1 Introduction

The complete solution of Linear-Quadratic Gaussian Dynamic Games (LQGDGs) has been a longstanding goal of the controls and games communities. That LQGDGs with a nonclassical information pattern can be problematic has been amply illustrated in Witsenhausen's seminal paper (Witsenhausen 1968)—see also Pachter and Pham (2014). Control theorists have traditionally emphasized control theoretic aspects and the backward induction/dynamic programming solution method, which however is not applicable to dynamic games with partial information—one notable exception notwithstanding, being the game with partial information that will be discussed herein. And game theorists have focused on information economics, that is, the role of information in games, but for the most part, discrete games. The state of affairs concerning dynamic games with partial information is not

M. Pachter (✉)
Department of Electrical and Computer Engineering, Air Force Institute of Technology, AFIT/ENG, 2950 Hobson Way, Bldg. 640, Wright-Patterson A.F.B., OH 45433, USA
e-mail: meir.pachter@afit.edu

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_8

satisfactory. In this respect, the situation is not much different now than it was in 1971 when Witsenhausen made a similar observation (Witsenhausen 1971). In this article a careful analysis of dynamic games with partial information is undertaken. We exclusively focus on LQGDGs, which are more readily amenable to analysis. Indeed, Linear-Quadratic Dynamic Games (LQDGs) with perfect information stand out as far as applications of the theory of dynamic games are concerned: a canonical instance of an application of the theory of LQDGs can be found in Ho et al. (1965) where it has been shown that its solution yields the Proportional Navigation (PN) guidance law which is universally used in Air-to-Air missiles. Furthermore, the theory of LQDGs has been successfully applied to the synthesis of H_∞ control laws (Basar and Bernhard 2008). The theory of LQDGs with perfect information has therefore received a great deal of attention (Basar and Olsder 1995; Engwerda 2005; Pachter and Pham 2010). In these works, the concepts of state, and state feedback, are emphasized and the solution method entails backward induction, a.k.a., Dynamic Programming (DP).

Concerning informational issues in LQGDGs: In previous work Radner (1962) and Pachter and Pham (2013) a static Linear-Quadratic Gaussian (LQG) team problem was addressed and a static "zero-sum" LQG game with partial information was analyzed in Pachter (2013). In this article a dynamic "zero-sum" LQG game, that is, a LQGDG, where the players have partial information, is addressed. The information pattern is as follows. The players' initial state information and their measurements are private information, but each player is able to observe the antagonist's past inputs: the protagonists' past controls is shared information. This information pattern has previously been discussed by Aoki (1973), and in the context of a team decision problem, this information pattern has also been discussed in Sandell and Athans (1974). However, Aoki (1973) took "a wrong turn": as so often happens in the literature of games with partial information, one is tempted to assume the players will try to second guess the opponents' private information, say, their measurements. The vicious cycle of second guessing the opponent's measurements leads to a mirror gallery like setting and to a dead end. This point is discussed in Sect. 8.4. Concerning reference Sandell and Athans (1974) where a decentralized dynamic team problem with a control-sharing information pattern is considered: It is argued that an infinite amount of information is contained in a real number, which, in theory, is correct. And since the control information is shared, then at least in a cooperative control/team setting, a player/agent could in principle encode in the controls information about to be sent to his partner his private information, for example, his measurements history. This is due to the fact that the controls which are about to be communicated can be modified slightly to encode the measurements information of the protagonists without significantly disturbing the control, and consequently, have a barely noticeable effect on the value of the game. One then falls back on the solution of the LQG cooperative control problem with a one-step-delay shared information pattern (Kurtaran and Sivan 1974). However, this scheme has no place in an antagonistic scenario, a.k.a., "zero-sum" LQGDG as discussed in our paper, and also does not properly model a decentralized control scenario. Moreover, this scheme totally depends on the players' ability of obtaining a noiseless observation of the broadcast control and as such, exhibits a lack of robustness to measurement error and is not a viable proposition. In the end, it is acknowledged in Sandell and Athans (1974) that the control-sharing information pattern leads to a stochastic control problem that is ill posed and it is stated that "the need for future work on this problem is obvious". Unfortunately, the analysis of LQGDGs with a control-sharing information pattern presented in Aoki (1973) and Sandell and Athans (1974) is patently incorrect. In this paper LQGDGs with a control-sharing information pattern are revisited. A careful analysis reveals that although this is a game with partial information, the control-sharing information pattern renders the game amenable to solution by the method of DP. It is shown that the solution of the LQGDG with a control-sharing information pattern is similar in structure to the solution of the LQG optimal control problem in so far as the principle of certainty equivalence/decomposition holds. A correct closed-form solution of a LQGDG with a control-sharing information pattern is obtained.

The paper is organized as follows. The LQGDG problem statement and the attendant control-sharing information pattern are presented in Sect. 8.2. The state estimation algorithm required for the solution of LQGDGs with a control-sharing information pattern is developed in Sect. 8.3. The analysis of Linear-Quadratic Gaussian Games with a control-sharing information pattern is anchored in Sect. 8.4 where the end-game is solved and the solution of the LQGDG with a control-sharing information pattern is obtained in Sect. 8.5 using the method of backward induction/DP. The results are summarized in Sect. 8.6, followed by concluding remarks in Sect. 8.7. For the sake of completeness, the solution of the baseline deterministic LQDG game with perfect information (Pachter and Pham 2010) is included in the Appendix. The somewhat lengthy exposition could perhaps be excused in light of Witsenhausen's observation when discussing LQG control (Witsenhausen 1971): "The most confused derivations of the correct results are also among the shortest".

8.2 Linear Quadratic Gaussian Dynamic Game

Two player Linear Quadratic Gaussian Dynamic Games (LQGDGs) are considered. The players are designated P and E, and the game is specified as follows.

Dynamics: Linear

$$x_{k+1} = A_k x_k + B_k u_k + C_k v_k + \Gamma_k w_k,\ x_0 \equiv x_0,\ \ k = 0, \ldots, N-1 \quad (8.1)$$

At decision time k the controls of players P and E are u_k and v_k, respectively. The process noise $w_k \sim \mathcal{N}(0, Q_p),\ \ k = 0, \ldots, N-1$. The planning horizon is N.

Measurements: Linear

The N measurements of player P are:

At time $k = 0$, $\bar{x}_0^{(P)}$— player P believes that the initial state

$$x_0 \sim \mathcal{N}(\bar{x}_0^{(P)}, P_0^{(P)}), \tag{8.2}$$

and thereafter he takes the measurements

$$z_{k+1}^{(P)} = H_{k+1}^{(P)} x_{k+1} + v_{k+1}^{(P)}, \ \ v_{k+1}^{(P)} \sim \mathcal{N}(0, R_m^{(P)}), \ k = 0, \ldots, N-2 \tag{8.3}$$

The N measurements of player E are:
At time $k = 0$, $\bar{x}_0^{(E)}$— player E believes that the initial state

$$x_0 \sim \mathcal{N}(\bar{x}_0^{E)}, P_0^{(E)}), \tag{8.4}$$

and thereafter he takes the measurements

$$z_{k+1}^{(E)} = H_{k+1}^{(E)} x_{k+1} + v_{k+1}^{(E)}, \ \ v_{k+1}^{(E)} \sim \mathcal{N}(0, R_m^{(E)}), \ k = 0, \ldots, N-2 \tag{8.5}$$

Cost/Payoff Function: Quadratic

We confine our attention to the antagonistic "zero-sum" game scenario where the respective P and E players strive to minimize and maximize the cost/payoff function

$$J = x_N^T Q_F x_N + \sum_{k=0}^{N-1} [x_{k+1}^T Q_{k+1} x_{k+1} + u_k^T R_k^{(P)} u_k - v_k^T R_k^{(E)} v_k] \rightarrow min_{\{u_k\}_{k=0}^{N-1}} max_{\{v_k\}_{k=0}^{N-1}} \tag{8.6}$$

Specifically, players P and E minimize and maximize their expected cost/payoff $E(J \mid \cdot)$, conditional on their private information. The expectation operator is liberally used in the dynamics game literature but oftentimes it is not clearly stated with respect to which random variables the expectation is calculated and on which random variables the expectation is conditional. This tends to mask the fact that what appear to be "zero-sum" games are in fact nonzero-sum games. Upon considering "zero-sum" games with partial information, the illusion is then created that a zero-sum game is considered. One then tends to rely on the uniqueness of the saddle point value and the interchangeability of non-unique optimal saddle point strategies in zero-sum games. This argument is flawed because, as previously discussed, in "zero-sum" games with partial information the P and E players calculate their respective cost and payoff conditional on their private information, as is correctly done in this paper; that's why I put the term zero-sum in quotation marks. Thus, although high powered mathematics is oftentimes used, serious conceptual errors make the "results" not applicable. Contrary to statements sometimes encountered in the literature, in "zero-sum" games with partial information one cannot look for a saddle point solution and the correct solution concept is a Nash equilibrium, that is, Person by Person Satisfactory (PBPS) solution. In this paper a unique Nash equilibrium is provided and the P and E players' value functions are calculated.

Information pattern

1. Public information
   (a) Problem parameters: A_k, B_k, C_k, $H_k^{(P)}$, $H_k^{(E)}$, Q_p, Q_k, Q_F, $R_k^{(P)}$, $R_k^{(E)}$, $R_m^{(P)}$, $R_m^{(E)}$.
   (b) Prior information: $P_0^{(P)}$, $P_0^{(E)}$.
2. Private information
   At decision time $k = 0$ the prior information of player P is $\bar{x}_0^{(P)}$.
   At decision time $k = 0$ the prior information of player E is $\bar{x}_0^{(E)}$.
   At decision time $1 \leq k \leq N-1$ the information of player P are his measurements $\bar{x}_0^{(P)}, z_1^{(P)}, \ldots, z_k^{(P)}$ and ownship control history $u_0, \ldots, u_{k-1}$.
   At decision time $1 \leq k \leq N-1$ the information of player E are his measurements $\bar{x}_0^{(E)}, z_1^{(E)}, \ldots, z_k^{(E)}$ and ownship control history $v_0, \ldots, v_{k-1}$.

Sufficient statistics

The sufficient statistics of player P at decision time $k = 0$: $x_0 \sim \mathcal{N}(\bar{x}_0^{(P)}, P_0^{(P)})$.
The sufficient statistics of player E at decision time $k = 0$: $x_0 \sim \mathcal{N}(\bar{x}_0^{(E)}, P_0^{(E)})$.
The sufficient statistics of player P at decision time $1 \leq k \leq N-1$: The p.d.f. $f_k^{(P)}(\cdot)$ of the physical state x_k, as calculated by player P using his private information.
The sufficient statistics of player E at decision time $1 \leq k \leq N-1$: The p.d.f. $f_k^{(E)}(\cdot)$ of the physical state x_k, as calculated by player E using his private information.

Remark. In the static LQGDG (Pachter 2013) where $N = 1$ the respective sufficient statistics of P and E are $\bar{x}_0^{(P)}$ and $\bar{x}_0^{(E)}$.

8.2.1 Problem Statement

The LQGDG (8.1)–(8.6) is considered and it is assumed that the players' information sets are augmented as follows: At decision time k, $k = 1, \ldots, N-1$, player P is endowed with the additional information regarding the control history $v_0, \ldots, v_{k-1}$ of player E. Thus, player P observed the past inputs $v_0, \ldots, v_{k-1}$ of player E. Similarly, at decision time k, $k = 1, \ldots, N-1$, player E is endowed with the additional information regarding the control history $u_0, \ldots, u_{k-1}$ of player P. Thus, player E observed the past inputs $u_0, \ldots, u_{k-1}$ of player P.

The information pattern considered herein is referred to as the *control-sharing* information pattern. The dynamics and the measurement equations are linear, and the cost/payoff function is quadratic, but the information pattern is not classical.

Strictly speaking, the information pattern is not partially nested because E's measurements, which he used to form his controls, are not known to P, and vice versa, P's measurements, which he used to form his controls, are not known to E. However, this is now a moot point because the information pattern is s.t. the control history of player E is known to player P, and vice versa, the control history of player P is known to player E. This, and the fact that player P and player E, each separately, perceive the initial state x_0 to be Gaussian, causes the state estimation problem faced by the players at decision time k to be Linear and Gaussian (LG). Hence, at decision time k, the knowledge of the complete control history $u_0, v_0, \ldots, u_{k-1}, v_{k-1}$ and their private measurement records makes it possible for both players to separately apply the linear Kalman filtering algorithm: based on his private measurements record, each player runs a Kalman filter using his measurements and the complete input history, and separately obtains an estimate of the state x_k—strictly speaking, the p.d.f. of x_k is separately obtained by each player. Thus, players P and E perceive the current state x_k to be Gaussian distributed. Having run their respective Kalman filters, at time k player P believes that the state

$$x_k \sim \mathscr{N}(\bar{x}_k^{(P)}, P_k^{(P)}) , \quad \forall\, k,\;\; N-1 \geq k \geq 0 \tag{8.7}$$

and player E believes that the state

$$x_k \sim \mathscr{N}(\bar{x}_k^{(E)}, P_k^{(E)}) , \quad \forall\, k,\;\; N-1 \geq k \geq 0 \tag{8.8}$$

but they are also aware that their state estimates are correlated—see Sect. 8.3.

Since the LQGDG (8.1)–(8.6) is LG, the P and E players' separately calculated sufficient statistics are given by Eqs. (8.7) and (8.8), and their controls will be determined by their optimal strategies according to $u_k^* = (\gamma_k^{(P)}(\bar{x}_k^{(P)}, P_k^{(P)}))^*$ and $v_k^* = (\gamma_k^{(E)}(\bar{x}_k^{(E)}, P_k^{(E)}))^*$. In fact, we shall show that in LQGDGs with a control-sharing information pattern the optimal strategies are of the form

$$u_k^* = (\gamma_k^{(P)}(\bar{x}_k^{(P)}))^* , \tag{8.9}$$

$$v_k^* = (\gamma_k^{(E)}(\bar{x}_k^{(E)}))^* , \quad \forall\, k,\;\; 0 \leq k \leq N-1 \tag{8.10}$$

and are linear.

8.3 Kalman Filtering

LQGDGs with a control-sharing information pattern are Linear-Gaussian (LG) and consequently at decision time k each player can separately calculate his estimate of the physical state x_k using a *linear* Kalman Filter (KF). Player P runs the KF

$$(\bar{x}_k^{(P)})^- = A\bar{x}_{k-1}^{(P)} + Bu_{k-1} + Cv_{k-1}, \ \ \bar{x}_0^{(P)} \equiv \bar{x}_0^{(P)} \tag{8.11}$$

$$(P_k^{(P)})^- = AP_{k-1}^{(P)}A^T + \Gamma Q_p \Gamma^T, \ \ P_0^{(P)} \equiv P_0^{(P)} \tag{8.12}$$

$$K_k^{(P)} = (P_k^{(P)})^-(H^{(P)})^T[H^{(P)}(P_k^{(P)})^-(H^{(P)})^T + R_m^{(P)}]^{-1} \tag{8.13}$$

$$\bar{x}_k^{(P)} = (\bar{x}_k^{(P)})^- + K_k^{(P)}[z_k^{(P)} - H^{(P)}(\bar{x}_k^{(P)})^-] \tag{8.14}$$

$$P_k^{(P)} = (I - K_k^{(P)}H^{(P)})(P_k^{(P)})^- \tag{8.15}$$

and so at decision time k player P obtains his estimate $\bar{x}_k^{(P)}$ of the state x_k. Similarly, player E runs the KF

$$(\bar{x}_k^{(E)})^- = A\bar{x}_{k-1}^{(E)} + Bu_{k-1} + Cv_{k-1}, \ \ \bar{x}_0^{(E)} \equiv \bar{x}_0^{(E)} \tag{8.16}$$

$$(P_k^{(E)})^- = AP_{k-1}^{(E)}A^T + \Gamma Q_p \Gamma^T, \ \ P_0^{(E)} \equiv P_0^{(E)} \tag{8.17}$$

$$K_k^{(E)} = (P_k^{(E)})^-(H^{(E)})^T[H^{(E)}(P_k^{(E)})^-(H^{(E)})^T + R_m^{(E)}]^{-1} \tag{8.18}$$

$$\bar{x}_k^{(E)} = (\bar{x}_k^{(E)})^- + K_k^{(E)}[z_k^{(E)} - H^{(E)}(\bar{x}_k^{(E)})^-] \tag{8.19}$$

$$P_k^{(E)} = (I - K_k^{(E)}H^{(E)})(P_k^{(E)})^- \tag{8.20}$$

and so at decision time k player E obtains his estimate $\bar{x}_k^{(E)}$ of the state x_k. The P and E players can calculate their respective state estimation error covariances and Kalman gains $P_k^{(P)}$, $K_{k+1}^{(P)}$, $P_k^{(E)}$ and $K_{k+1}^{(E)}$ ahead of time and off line.

In LQGDGs with a control-sharing information pattern the players' sufficient statistic is their state estimate; the latter is the argument of their strategy functions (8.9) and (8.10). Hence, in the process of countering E's action, P must compute the *statistics* of E's state estimate $\bar{x}_k^{(E)}$, and, vice versa, while planning his move, E must compute the *statistics* of P's state estimate $\bar{x}_k^{(P)}$. Momentarily assume the point of view of player P: As far as P is concerned, the unknown to him state estimate of player E at time k, $\bar{x}_k^{(E)}$, is a random variable (and consequently E's input at time k is a random variable). Similarly, player E will consider the unknown to him state estimate of player P at time k, $\bar{x}_k^{(P)}$, to be a random variable (and consequently P's input at time k is a random variable). Hence, in the LQGDG with a control-sharing information pattern, at time k player P will estimate E's state estimate $\bar{x}_k^{(E)}$ using his calculated ownship state estimate $\bar{x}_k^{(P)}$, and vice versa, player E will estimate P's state estimate $\bar{x}_k^{(P)}$ using his calculated ownship state estimate $\bar{x}_k^{(E)}$. Thus, in the LQGDG with a control-sharing information pattern and with his state estimate $\bar{x}_k^{(P)}$ at time k in hand, player P calculates the statistics of E's state estimate $\bar{x}_k^{(E)}$, conditional on the public and private information available to him at time k. Similarly, having obtained at time k his state estimate $\bar{x}_k^{(E)}$, player E calculates the statistics of the state estimate $\bar{x}_k^{(P)}$ of player P, conditional on the public and private information available to him at time k. Let's start at decision time $k = 0$.

Player P models his measurement/estimate $\overline{x}_0^{(P)}$ of the initial state x_0 as

$$\overline{x}_0^{(P)} = x_0 + e_0^{(P)} , \tag{8.21}$$

where x_0 is the true physical state and $e_0^{(P)}$ is player P's measurement/estimation error, whose statistics, in view of Eq. (8.2), are $e_0^{(P)} \sim \mathcal{N}(0, P_0^{(P)})$. In addition, player P models player E's measurement $\overline{x}_0^{(E)}$ of the initial state x_0 as

$$\overline{x}_0^{(E)} = x_0 + e_0^{(E)} , \tag{8.22}$$

where, as before, x_0 is the true physical state and $e_0^{(E)}$ is player E's measurement/estimation error, whose statistics, which are known to P—see Eq. (8.4)—are $e_0^{(E)} \sim \mathcal{N}(0, P_0^{(E)})$. The Gaussian random variables $e_0^{(P)}$ and $e_0^{(E)}$ are *independent*—by hypothesis. From player P's point of view, $\overline{x}_0^{(P)}$ is known but $\overline{x}_0^{(E)}$ is a random variable. Subtracting Eq. (8.21) from Eq. (8.22), at time $k = 0$ player P concludes that as far as he is concerned, player E's measurement upon which he will decide, according to Eq. (8.10), on his optimal control v_0^*, is the random variable

$$\overline{x}_0^{(E)} = \overline{x}_0^{(P)} + e_0^{(E)} - e_0^{(P)} , \tag{8.23}$$

In other words, as far as P is concerned, E's estimate $\overline{x}_0^{(E)}$ of the initial state x_0 is the Gaussian random variable

$$\overline{x}_0^{(E)} \sim \mathcal{N}(\overline{x}_0^{(P)}, P_0^{(P)} + P_0^{(E)}) \tag{8.24}$$

Thus, player P has used his measurement/private information $\overline{x}_0^{(P)}$ and the public information $P_0^{(P)}$ and $P_0^{(E)}$ to calculate the statistics of the sufficient statistic $\overline{x}_0^{(E)}$ of player E, which is the argument of E's strategy function $\gamma_0^{(E)}(\cdot)$; the latter, along with P's control u_0, will feature in player's P cost functional. Similarly, as far as player E is concerned, at time $k = 0$ the statistics of the sufficient statistic $\overline{x}_0^{(P)}$ of player P are

$$\overline{x}_0^{(P)} \sim \mathcal{N}(\overline{x}_0^{(E)}, P_0^{(P)} + P_0^{(E)}) \tag{8.25}$$

Similar to the case where $k = 0$, as far as player P is concerned the state estimate of player E at decision time $k \geq 1$ is the random variable

$$\overline{x}_k^{(E)} = \overline{x}_k^{(P)} + e_k^{(E)} - e_k^{(P)},$$

that is, at decision time k player P believes that the state estimate $\overline{x}_k^{(E)}$ of player E is

$$\overline{x}_k^{(E)} \sim \mathcal{N}(\overline{x}_k^{(P)}, P_k^{(E,P)}) \tag{8.26}$$

where the covariance matrix

$$P_k^{(E,P)} \equiv E(\ (e_k^{(E)} - e_k^{(P)})(e_k^{(E)} - e_k^{(P)})^T\ )$$
$$= P_k^{(P)} + P_k^{(E)} - E(\ e_k^{(P)}(e_k^{(E)})^T\ ) - (E(\ e_k^{(P)}(e_k^{(E)})^T\ ))^T$$

At the decision time instants $k = 1, \ldots, N-1$ the P and E players' respective state estimation errors $e_k^{(P)}$ and $e_k^{(E)}$ are now correlated—this is caused by the process dynamics being driven in part by process noise.

Similarly, as far as he is concerned, player E believes that at decision time k the state estimate $\bar{x}_k^{(P)}$ of player P is the random variable

$$\bar{x}_k^{(P)} \sim \mathcal{N}(\bar{x}_k^{(E)}, P_k^{(E,P)}) \tag{8.27}$$

Concerning decision time $k \geq 1$: Let the covariance matrix

$$\tilde{P}_k^{(E,P)} \equiv E(\ e_k^{(P)}(e_k^{(E)})^T\ ) \tag{8.28}$$

It can be shown that the recursion for the correlation matrix $\tilde{P}_k^{(E,P)}$ is

$$\tilde{P}_{k+1}^{(E,P)} = (I - K_{k+1}^{(P)}H^{(P)})(A\tilde{P}_k^{(P,E)}A^T + \Gamma Q_p \Gamma^T)(I - K_{k+1}^{(E)}H^{(E)})^T,\ \tilde{P}_0^{(P,E)} = 0,$$
$$k = 0, \ldots, N-1 \tag{8.29}$$

In summary, at decision time $k = 0, \ldots, N-1$ player P believes that the statistics of E's estimate $\bar{x}_k^{(E)}$ of the state x_k are given by Eq. (8.26) and player E believes that the statistics of P's estimate $\bar{x}_k^{(P)}$ of the state x_k are given by Eq. (8.27) where

$$P_k^{(E,P)} = P_k^{(P)} + P_k^{(E)} - \tilde{P}_k^{(E,P)} - (\tilde{P}_k^{(E,P)})^T$$

The KF covariance matrices $P_k^{(P)}$, $P_k^{(E)}$ and $\tilde{P}_k^{(E,P)}$ are calculated ahead of time by solving the respective recursions (8.12), (8.13), (8.15); (8.17), (8.18), (8.20); and (8.29).

Finally, since in LQGDGs with a control-sharing information pattern the sufficient statistic is the players' state estimate, then upon employing the method of Dynamic Programming, at decision time k player P must project ahead the *estimate* of the physical state x_{k+1} that the Kalman filtering algorithm will provide at time $k+1$. It can be shown that at time k player P believes that the future state x_{k+1} at time $k+1$ will be the Gaussian random variable

$$\bar{x}_{k+1}^{(P)} = A\bar{x}_k^{(P)} + Bu_k + C\gamma_k^{(E)}(\bar{x}_k^{(P)} + e_k^{(P)} - e_k^{(E)}) + K_{k+1}^{(P)}(H^{(P)}\Gamma w_k$$
$$+ v_{k+1}^{(P)} - H^{(P)}Ae_k^{(P)}) \tag{8.30}$$

Similarly, at decision time k player E's estimate of the state x_{k+1} at time $k+1$ will be the Gaussian random variable

$$\overline{x}_{k+1}^{(E)} = A\overline{x}_k^{(E)} + B\gamma_k^{(P)}(\overline{x}_k^{(E)} + e_k^{(E)} - e_k^{(P)}) + Cv_k + K_{k+1}^{(E)}(H^{(E)}\Gamma w_k + v_{k+1}^{(E)} - H^{(E)}Ae_k^{(E)}) \quad (8.31)$$

8.4 End Game

In the best tradition of backward induction/Dynamic Programming, the terminal stage of the game, namely, the players' decision time $k = N-1$ is analyzed first. In the end game the cost/payoff function is

$$\begin{aligned} J_{N-1}(u_{N-1}, v_{N-1}; x_{N-1}) &= x_N^T Q_F x_N + x_N^T Q x_N + u_{N-1}^T R_u u_{N-1} - v_{N-1}^T R_v v_{N-1} \\ &= x_N^T (Q + Q_F) x_N + u_{N-1}^T R_u u_{N-1} - v_{N-1}^T R_v v_{N-1} \end{aligned}$$

It is convenient to momentarily set $Q_F := Q + Q_F$ whereupon the terminal cost/payoff

$$J_{N-1}(u_{N-1}, v_{N-1}; x_{N-1}) = x_N^T Q_F x_N + u_{N-1}^T R_u u_{N-1} - v_{N-1}^T R_v v_{N-1} \quad (8.32)$$

The players' sufficient statistics in this LG game are the expectation of the physical state and the covariance of the state's estimation error: having run their respective Kalman filters during the time interval $[1, N-1]$, at decision time $N-1$ the information available to player P is $(\overline{x}_{N-1}^{(P)}, P_{N-1}^{(P)})$ and the information of player E is $(\overline{x}_{N-1}^{(E)}, P_{N-1}^{(E)})$. In other words, at decision time $N-1$ player P believes the physical state x_{N-1} to be $x_{N-1} \sim \mathcal{N}(\overline{x}_{N-1}^{(P)}, P_{N-1}^{(P)})$ whereas player E believes the physical state x_{N-1} to be specified as $x_{N-1} \sim \mathcal{N}(\overline{x}_{N-1}^{(E)}, P_{N-1}^{(E)})$. This is tantamount to stipulating that players P and E took separate measurements of the state x_{N-1}. The quality of the players' "instruments" used to take the measurements and also the degree of correlation of the players' measurement errors is public knowledge—we refer to the measurement error covariances $P_{N-1}^{(E)}$, $P_{N-1}^{(E)}$ and $\tilde{P}_{N-1}^{(E,P)}$. At the same time, the recorded measurements $\overline{x}_{N-1}^{(P)}$and $\overline{x}_{N-1}^{(E)}$ are the private information of the respective players P and E: the "measurement" $\overline{x}_{N-1}^{(E)}$ recorded by player E is his private information and is not shared with player P. Thus, player P has partial information. Similarly, the "measurement" $\overline{x}_{N-1}^{(P)}$ recorded by player P is his private information and is not shared with player E, so also player E has partial information.

To gain a better appreciation of the informational issues in games with partial information, it is instructive to briefly digress and employ an "appealing" approach which is familiar to workers in deterministic control and which, unfortunately, is an approach sometimes employed in stochastic games. We now intentionally "take a wrong turn" which quickly leads us to a dead end. A correct analysis of the informational situation at hand follows.

Consider the following flawed argument: At time $N-1$ the state information available to player P is $x_{N-1} \sim \mathcal{N}(\bar{x}^{(P)}_{N-1}, P^{(P)}_{N-1})$ and thus Player P calculates the expectation of his cost function

$$\begin{aligned}\bar{J}^{(P)}_{N-1}(u_{N-1}, v_{N-1}; \bar{x}^{(P)}_{N-1}, P^{(P)}_{N-1}) &\equiv E_{x_{N-1}}\left(J(u_{N-1}, v_{N-1}; x_{N-1}) \mid \bar{x}^{(P)}_{N-1}, P^{(P)}_{N-1}\right)\\ &= (\bar{x}^{(P)})^T_{N-1} A^T Q_F A \bar{x}^{(P)}_{N-1} + Trace(A^T Q_F A P^{(P)}_{N-1})\\ &+ u^T_{N-1}(R_u + B^T Q_F B) u_{N-1} - v^T_{N-1}(R_v - C^T Q_F C) v_{N-1}\\ &+ 2u^T_{N-1} B^T Q_F A \bar{x}^{(P)}_{N-1} + 2v^T_{N-1} C^T Q_F A \bar{x}^{(P)}_{N-1}\\ &+ 2u^T_{N-1} B^T Q_F C v_{N-1} + Trace(\Gamma^T Q_F \Gamma Q_p) \qquad (8.33)\end{aligned}$$

At the same time the state information available to player E is $x_{N-1} \sim \mathcal{N}(\bar{x}^{(E)}_{N-1}, P^{(E)}_{N-1})$ and Player E calculates the expectation of his payoff function

$$\begin{aligned}\bar{J}^{(E)}_{N-1}(u_{N-1}, v_{N-1}; \bar{x}^{(E)}_{N-1}, P^{(E)}_{N-1}) &\equiv E_{x_{N-1}}\left(J(u_{N-1}, v_{N-1}; x_{N-1}) \mid \bar{x}^{(E)}_{N-1}, P^{(E)}_{N-1}\right)\\ &= (\bar{x}^{(E)})^T_{N-1} A^T Q_F A \bar{x}^{(E)}_{N-1} + Trace(A^T Q_F A P^{(E)}_{N-1})\\ &+ u^T_{N-1}(R_u + B^T Q_F B) u_{N-1} - v^T_{N-1}(R_v - C^T Q_F C) v_{N-1}\\ &+ 2u^T_{N-1} B^T Q_F A \bar{x}^{(E)}_{N-1} + 2v^T_{N-1} C^T Q_F A \bar{x}^{(E)}_{N-1}\\ &+ 2u^T_{N-1} B^T Q_F C v_{N-1} + Trace(\Gamma^T Q_F \Gamma Q_p) \qquad (8.34)\end{aligned}$$

Now Player P's optimization, that is, the differentiation of his *deterministic* cost function (8.33), yields the relationship

$$(R_u + B^T Q_F B) u_{N-1} + B^T Q_F C v_{N-1} = -B^T Q_F A \bar{x}^{(P)}_{N-1} \qquad (8.35)$$

and Player E's optimization, that is, the differentiation of his *deterministic* payoff function (8.34), yields the relationship

$$C^T Q_F B u_{N-1} - (R_v - C^T Q_F C) v_{N-1} = -C^T Q_F A \bar{x}^{(E)}_{N-1} \qquad (8.36)$$

Have obtained two equations in the players' optimal *controls*, namely, the two unknowns u^*_{N-1} and v^*_{N-1}, which players P and E must *separately* solve in order to calculate their respective optimal controls. However player P cannot solve the set of two equations (8.35) and (8.36) because he does *not* know the "measurement" $\bar{x}^{(E)}_{N-1}$ of E, and player E cannot solve this set of two equations because he does *not* know the "measurement" $\bar{x}^{(P)}_{N-1}$ of P—both players have reached a dead end and it would appear that all that's left to do is try to guess and outguess the opponent's "measurement". This state of affairs is caused by the players having partial information. This approach brings on the much maligned infinite regress in reciprocal reasoning! Unfortunately, this flawed approach is not foreign to the literature on dynamic stochastic games and it leads to erroneous "results"—see Aoki (1973) where, using this flawed argument, the LQGDG with a shared-control information pattern was "solved" and complicated "strategies" were computed.

We now change course and undertake a correct analysis of our LQGDG with a shared-control information pattern. To this end, it is imperative that one thinks in *strategic* terms. The strategies available to player P are mappings $f : R^n \to R^{m_u}$ from his information set into his actions set; thus, the action of player P is $u_{N-1} = f(\overline{x}^{(P)}_{N-1}, P^{(P)}_{N-1})$. Similarly, the strategies available to player E are mappings $g : R^n \to R^{m_v}$ from his information set into his actions set—thus, the action of player E is $v_{N-1} = g(\overline{x}^{(E)}_{N-1}, P^{(E)}_{N-1})$. However, we'll show in the sequel that it suffices to consider P and E strategies of the form (8.9) and (8.10), respectively.

It is now important to realize that from player P's vantage point, the action v_{N-1} of player E is a *random* variable. This is so because as far as player P is concerned the measurement $\overline{x}^{(E)}_{N-1}$ of player E used in (8.10) to form his control v_{N-1} is a random variable. Similarly, from player E's vantage point, the action u_{N-1} of player P is also a function of a random variable, $\overline{x}^{(P)}_{N-1}$.

Consider the decision process of player P whose private information is $\overline{x}^{(P)}_{N-1}$. He operates against the *strategy* $g(\cdot)$ of player E. Therefore, from player P's perspective, the random variables at work are x_{N-1} *and* $\overline{x}^{(E)}_{N-1}$. At decision time $k = N-1$ player P is confronted with a *stochastic* optimization problem and he calculates the *expectation* of the cost function (8.32), conditional on his private information $\overline{x}^{(P)}_{N-1}$,

$$\overline{J}^{(P)}(u_{N-1}, g(\cdot); \overline{x}^{(P)}_{N-1}) \equiv E_{x_{N-1}, \overline{x}^{(E)}_{N-1}}\left(J(u_{N-1}, g(\overline{x}^{(E)}_{N-1}); x_{N-1}) \mid \overline{x}^{(P)}_{N-1}\right) \tag{8.37}$$

By correctly using in the calculation of his expected cost (8.37) player's E *strategy function* $g(\overline{x}^{(E)}_{N-1})$ rather than, as before, player E's *control* v_{N-1}, player P has eliminated the possibility of an infinite regress in reciprocal reasoning. This is so because P now has all the information to be able, in principle, to calculate the said expectation. Thus, player P calculates his expected cost

$$\begin{aligned}\overline{J}^{(P)}(u_{N-1}, g(\cdot); \overline{x}^{(P)}_{N-1}) = {} & (\overline{x}^{(P)}_{N-1})^T A^T Q_F A \overline{x}^{(P)}_{N-1} + Trace(A^T Q_F A P^{(P)}_{N-1}) \\ & + u^T_{N-1}(R_u + B^T Q_F B) u_{N-1} + 2u^T_{N-1} B^T Q_F A \overline{x}^{(P)}_{N-1} \\ & + 2E_{x_{N-1}, \overline{x}^{(E)}_{N-1}}\left(g^T(\overline{x}^{(E)}_{N-1}) C^T Q_F A x_{N-1} \mid \overline{x}^{(P)}_{N-1}\right) \\ & - E_{\overline{x}^{(E)}_{N-1}}\left(g^T(\overline{x}^{(E)}_{N-1})(R_v - C^T Q_F C) g(\overline{x}^{(E)}_{N-1}) \mid \overline{x}^{(P)}_{N-1}\right) \\ & + 2u^T_{N-1} B^T Q_F C E_{\overline{x}^{(E)}_{N-1}}\left(g(\overline{x}^{(E)}_{N-1}) \mid \overline{x}^{(P)}_{N-1}\right) \\ & + Trace(\Gamma^T Q_F \Gamma Q_p)\end{aligned} \tag{8.38}$$

Player P calculates the expectations with respect to the random variable $\bar{x}_{N-1}^{(E)}$ which features in Eq. (8.38), cognizant that it is $\bar{x}_{N-1}^{(E)} \sim \mathscr{N}(\bar{x}_{N-1}^{(P)}, P_{N-1}^{(E,P)})$. In this game with partial information, player P is using his measurement/private information $\bar{x}_{N-1}^{(P)}$ and the public information to estimate the sufficient statistic $\bar{x}_{N-1}^{(E)}$ of player E, which is the argument of E's strategy function $g(\cdot)$; the latter features in player's P cost functional (8.38) and thus enters the calculation of P's cost.

The careful analysis of the optimization problem at hand leads to a Fredholm equation of the second kind of the convolution type with a kernel which is a Gaussian function; the unknown functions are the players' optimal strategies. Taking the point of view of player E yields a similar Fredholm integral equation in the players' optimal strategies. The solution of the set of two Fredholm equations yields the optimal strategies of players P and E. The optimal strategies turn out to be linear after all! The reader is referred to reference Pachter (2013) for the complete derivation.

8.5 Dynamic Programming

We consider the LQGDG (8.1)–(8.6) with a control-sharing information pattern as in Aoki (1973). The planning horizon $N \geq 2$.

8.5.1 Sufficient Statistics

The initial state information and the measurements of players P and E are their private information but their past controls are shared information. Even though the players have partial information because the initial state information and their measurements are not shared, from the point of view of both players P and E, the control system is nevertheless Linear Gaussian (LG). This is so because at decision time k their respective adversary's information state components $v_0, \ldots, v_{k-1}$ and $u_0, \ldots, u_{k-1}$ are *not* random variables with unknown p.d.f.s but are known to the players: The LQGDG with a control-sharing information pattern is LG and therefore the conditions for the P-player's information state to be Gaussian hold and at decision time k the sufficient statistics of P and E are $\bar{x}_k^{(P)}$ and $\bar{x}_k^{(E)}$, respectively. Furthermore, as far as player P is concerned, at time k the sufficient statistic $\bar{x}_k^{(E)}$ of player E is the random variable $\bar{x}_k^{(E)} \sim \mathscr{N}(\bar{x}_k^{(P)}, P_k^{(E,P)})$ and he uses this information in the calculation of his cost-to-go/value function at time k. Similarly, player E considers the sufficient statistic $\bar{x}_k^{(P)}$ of player P to be $\bar{x}_k^{(P)} \sim \mathscr{N}(\bar{x}_k^{(E)}, P_k^{(E,P)})$ and player E uses this information in the calculation of his cost-to-go/value function at time k.

8.5.2 Analysis

The analysis is along the lines of the analysis of the static LQG game with partial information (Pachter 2013) and the analysis of the end game in Sect. 8.4 where $k = N - 1$. We shall require

Proposition 1. *The value functions of players P and E are quadratic in their respective sufficient statistics* $\bar{x}_k^{(P)}$ *and* $\bar{x}_k^{(E)}$*, that is*

$$V_k^{(P)}(\bar{x}_k^{(P)}, P_k^{(P)}; P_k^{(E)}, \tilde{P}_k^{(E,P)}) = (\bar{x}_k^{(P)})^T \Pi_k \bar{x}_k^{(P)} + c_k^{(P)}(P_k^{(P)}; P_k^{(E)}, \tilde{P}_k^{(E,P)}), \;\; k = 0, 1, \ldots, N-1,$$
$$V_k^{(E)}(\bar{x}_k^{(E)}, P_k^{(E)}; P_k^{(P)}, \tilde{P}_k^{(E,P)}) = (\bar{x}_k^{(E)})^T \Pi_k \bar{x}_k^{(E)} + c_k^{(E)}(P_k^{(E)}; P_k^{(P)}, \tilde{P}_k^{(E,P)}), \;\; k = 0, 1, \ldots, N-1$$

where

Π_k are $n \times n$ real symmetric matrices and the scalars $c_k^{(P)}, c_k^{(E)} \in R^1, \; k = 0, \ldots, N.$

□

Similar to the *correct* approach outlined in Sect. 8.4 we calculate the value functions by taking the expectations over the relevant random variables.

$$\begin{aligned}
V_k^{(P)}(\bar{x}_k^{(P)}, P_k^{(P)}; P_k^{(E)}, \tilde{P}_k^{(E,P)}) = \; & min_{u_k}\{u_k^T[R_u + B^T(Q + \Pi_{k+1})B]u_k \\
& +2u_k^T B^T(Q + \Pi_{k+1})(A\bar{x}_k^{(P)} \\
& + CE_{\tilde{w}}\,(\gamma_k^{(E)}(\bar{x}_k^{(P)} + \tilde{w})))\} + (\bar{x}_k^{(P)})^T A^T(Q + \Pi_{k+1})A\bar{x}_k^{(P)} \\
& - E_{\tilde{w}}\,((\gamma_k^{(E)}(\bar{x}_k^{(P)} + \tilde{w}))^T \\
& [R_v - C^T(Q + \Pi_{k+1})C]\gamma_k^{(E)}(\bar{x}_k^{(P)} + \tilde{w})) \\
& + 2E_{\tilde{w}}\,((\gamma_k^{(E)}(\bar{x}_k^{(P)} + \tilde{w}))^T)C^T(Q + \Pi_{k+1})A\bar{x}_k^{(P)} \\
& - 2E_{e_k^{(E)}, e_k^{(P)}}\,((\gamma_k^{(E)}(\bar{x}_k^{(P)} + e_k^{(E)} - e_k^{(P)}))^T C^T(Q \\
& + \Pi_{k+1}K_{k+1}^{(P)}H^{(P)})Ae_k^{(P)}) + Trace(A^T QAP_k^{(P)}) \\
& + Trace(\Gamma^T Q\Gamma Q_p) + Trace((K_{k+1}^{(P)})^T \Pi_{k+1}K_{k+1}^{(P)}R_m^{(P)}) \\
& + Trace(\Gamma^T(H^{(P)})^T(K_{k+1}^{(P)})^T \Pi_{k+1}K_{k+1}^{(P)}H^{(P)}\Gamma Q_p) \\
& + Trace(A^T(H^{(P)})^T(K_{k+1}^{(P)})^T \Pi_{k+1}K_{k+1}^{(P)}H^{(P)}AP_k^{(P)}) \\
& + c_{k+1}^{(P)}(P_{k+1}^{(P)}; P_{k+1}^{(E)}, \tilde{P}_{k+1}^{(E,P)}),
\end{aligned} \tag{8.39}$$

$$\begin{aligned}
V_k^{(E)}(\overline{x}_k^{(E)}, P_k^{(E)}; P_k^{(P)}, \tilde{P}_k^{(E,P)}) = & \, max_{v_k}\{-v_k^T[R_v - C^T(Q + \Pi_{k+1})C]v_k \\
& +2v_k^T C^T(Q + \Pi_{k+1})(A\overline{x}_k^{(E)} \\
& + BE_{\tilde{w}}\,(\gamma_k^{(P)}(\overline{x}_k^{(E)} - \tilde{w})))\} + (\overline{x}_k^{(E)})^T A^T(Q + \Pi_{k+1})A\overline{x}_k^{(E)} \\
& + E_{\tilde{w}}\,((\gamma_k^{(P)}(\overline{x}_k^{(E)} - \tilde{w}))^T \\
& [R_u + B^T(Q + \Pi_{k+1})B]\gamma_k^{(P)}(\overline{x}_k^{(E)} - \tilde{w})) \\
& + 2E_{\tilde{w}}\,((\gamma_k^{(P)}(\overline{x}_k^{(E)} - \tilde{w}))^T)B^T(Q + \Pi_{k+1})A\overline{x}_k^{(E)} \\
& - 2E_{e_k^{(E)}, e_k^{(P)}}\,((\gamma_k^{(P)}(\overline{x}_k^{(E)} - e_k^{(E)} + e_k^{(P)}))^T B^T(Q \\
& + \Pi_{k+1}K_{k+1}^{(E)}H^{(E)})Ae_k^{(E)}) + Trace(A^T QAP_k^{(E)}) \\
& + Trace(\Gamma^T Q\Gamma Q_p) + Trace((K_{k+1}^{(E)})^T \Pi_{k+1}K_{k+1}^{(E)}R_m^{(E)}) \\
& + Trace(\Gamma^T(H^{(E)})^T(K_{k+1}^{(E)})^T P_{k+1}K_{k+1}^{(E)}H^{(E)}\Gamma Q_p) \\
& + Trace(A^T(H^{(E)})^T(K_{k+1}^{(E)})^T \Pi_{k+1}K_{k+1}^{(E)}H^{(E)}AP_k^{(E)}) \\
& + c_{k+1}^{(E)}(P_{k+1}^{(P)}; P_{k+1}^{(E)}, \tilde{P}_{k+1}^{(E,P)}) \qquad (8.40)
\end{aligned}$$

where the random variable $\tilde{w} \equiv e_k^{(P)} - e_k^{(E)} \sim \mathcal{N}(0, P_k^{(E,P)})$.

8.5.3 Optimization

Consider the minimization problem faced by P at decision time $0 \leq k \leq N-2$: Differentiating the RHS of Eq. (8.39) in his *control* u_k he obtains the optimality condition

$$u_k^* = -[R_u + B^T(Q + \Pi_{k+1})B]^{-1}B^T(Q + \Pi_{k+1})(A\overline{x}_k^{(P)} + CE_{\tilde{w}}(\,(\gamma_k^{(E)}(\overline{x}_k^{(P)} + \tilde{w})\,)),$$
$$k = 0, 1, \ldots, N-1 \qquad (8.41)$$

and similarly, upon differentiating the RHS of Eq. (8.40) in v_k player E obtains

$$v_k^* = [R_v - C^T(Q + \Pi_{k+1})C]^{-1}C^T(Q + \Pi_{k+1})(A\overline{x}_k^{(E)} + BE_{\tilde{w}}(\,(\gamma_k^{(P)}(\overline{x}_k^{(E)} - \tilde{w})\,)),$$
$$k = 0, 1, \ldots, N-1 \qquad (8.42)$$

Player P has obtained an expression for his optimal control u_k^* where $\overline{x}_k^{(E)}$ does not feature and u_k^* is a function of the parameter $\overline{x}_k^{(P)}$ only. However, the strategy function $\gamma_k^{(E)}(\cdot)$ of player E features in this equation. Indeed, the strategic relationship holds

$$(\gamma_k^{(P)}(\bar{x}_k^{(P)}))^* = -[R_u + B^T(Q+\Pi_{k+1})B]^{-1}B^T(Q+\Pi_{k+1})(A\bar{x}_k^{(P)} + CE_{\tilde{w}}(\,((\gamma_k^{(E)}(\bar{x}_k^{(P)}+\tilde{w}))^*\,)), \quad k=0,1,\ldots,N-1 \quad (8.43)$$

We have obtained an expression for P's optimal strategy function $(\gamma_k^{(P)}(\bar{x}_k^{(P)}))^*$ in terms of the strategy $\gamma_k^{(E)}(\cdot)$ of player E. Payer P obtained a linear relationship which directly ties together the as yet unknown optimal strategies $(\gamma_k^{(P)}(\bar{x}_k^{(P)}))^*$ and $(\gamma_k^{(E)}(\bar{x}_k^{(E)}))^*$ of players P and E. Similarly, also player E obtains a linear relationship among the players' optimal *strategies*:

$$(\gamma_k^{(E)}(\bar{x}_k^{(E)}))^* = [R_v - C^T(Q+\Pi_{k+1})C]^{-1}C^T(Q+\Pi_{k+1})(A\bar{x}_k^{(E)} + BE_{\tilde{w}}(\,((\gamma_k^{(P)}(\bar{x}_k^{(E)}-\tilde{w}))^*\,)), \quad k=0,1,\ldots,N-1 \quad (8.44)$$

Equations (8.43) and (8.44) constitute a linear system of Fredholm integral equations of the second kind in the players' optimal *strategies* $(\gamma_k^{(P)}(\bar{x}_k^{(P)}))^*$ and $(\gamma_k^{(E)}(\bar{x}_k^{(E)}))^*$. Similar to the analysis in reference Pachter (2013), the solution of the linear system of Fredholm integral equations of the second kind, Eqs. (8.43) and (8.44), yields the optimal strategies which are linear in the players' sufficient statistics, namely

$$\gamma_k^{(P)}(\bar{x}_k^{(P)}) = F_k^{(P)}\cdot\bar{x}_k^{(P)}, \quad \gamma_k^{(E)}(\bar{x}_k^{(E)}) = F_k^{(E)}\cdot\bar{x}_k^{(E)}$$

and the formulae for the optimal gains

$$(F_k^{(P)})^* = -S_B^{-1}(Q+\Pi_{k+1})B^T(Q+\Pi_{k+1})\{I + C[R_v - C^T(Q+\Pi_{k+1})C]^{-1}C^T(Q + \Pi_{k+1})\}A, \qquad k=0,\ldots,N-1 \quad (8.45)$$

$$(F_k^{(E)})^* = -S_C^{-1}(Q+\Pi_{k+1})C^T(Q+\Pi_{k+1})\{I - B[R_u + B^T(Q+\Pi_{k+1})B]^{-1}B^T(Q + \Pi_{k+1})\}A, \qquad k=0,\ldots,N-1 \quad (8.46)$$

The control system is Linear - Gaussian (LG) and therefore the players' information states are Gaussian, time consistency in this dynamic game is provided by the application of the method of Dynamic Programming (DP) where the DP state is the information state, and, by construction, the strategies are Person-By-Person-Satisfactory (PBPS), so in the LQGDG with a control-sharing information pattern, a Nash equilibrium is obtained—as was also the case in the static LQG game with partial information (Pachter 2013).

8.5.4 Value Functions

The parameters which specify the statistics of the random variables in the LQGDG do not feature in the formulae (8.45) and (8.46) for the players' optimal strategies and consequently an inspection of the DP equations (8.39) and (8.40) tells us that the matrices Π_k won't be a function of the said parameters; in other words, the matrices Π_k are exclusively determined by the deterministic plant's parameters A, B, C, Q, Q_F, $R_c^{(P)}$ and $R_c^{(E)}$. Hence $\Pi_k = P_k$, where P_k is the solution of the Riccati equation (8.57) derived for the *deterministic* LQDG discussed in the Appendix. Upon defining $P_k := P_k + Q$, the optimal gains correspond to the optimal gains in the deterministic LQDG, Eqs. (8.61) and (8.62) in the Appendix and the players' optimal gains are

$$(F_k^{(P)})^* = -S_B^{-1}(P_{k+1}+Q)B^T(P_{k+1}+Q)\{I + C[R_v - C^T(P_{k+1}+Q)C]^{-1}C^T(P_{k+1} +Q)\}A \tag{8.47}$$

$$(F_k^{(E)})^* = -S_C^{-1}(P_{k+1}+Q)C^T(P_{k+1}+Q)\{I - B[R_u + B^T(P_{k+1}+Q)B]^{-1}B^T(P_{k+1} +Q)\}A \tag{8.48}$$

The recursions for the scalars $c_k^{(P)}$ and $c_k^{(E)}$ are obtained from the respective DP equations (8.39) and (8.40):

$$\begin{aligned} c_k^{(P)} &= c_{k+1}^{(P)} \\ &+ 2\,Trace((F_k^{(P)}))^{*T} C^T (Q + P_{k+1} K_{k+1}^{(P)} H^{(P)}) A (P_k^{(P)} - \tilde{P}_k^{(P,E)}) \\ &+ Trace(A^T Q A P_k^{(P)}) + Trace(\Gamma^T Q \Gamma Q_p) \\ &+ Trace(\Gamma^T (H^{(P)})^T (K_{k+1}^{(P)})^T P_{k+1} K_{k+1}^{(P)} H^{(P)} \Gamma Q_p) \\ &+ Trace((K_{k+1}^{(P)})^T P_{k+1} K_{k+1}^{(P)} R_m^{(P)}) \\ &+ Trace(A^T (H^{(P)})^T (K_{k+1}^{(P)})^T P_{k+1} K_{k+1}^{(P)} H^{(P)} A P_k^{(P)}),\ k = N-2,\ldots,0 \end{aligned} \tag{8.49}$$

and for $k = N-1$ we use the end-game equation

$$\begin{aligned} c_{N-1}^{(P)}(P_{N-1}^{(P)}; P_{N-1}^{(E)}, \tilde{P}_{N-1}^{(P,E)}) = Trace(\, & A^T Q_F A P_{N-1}^{(P)} + 2(P_{N-1}^{(P)} \\ & - \tilde{P}_{N-1}^{(P,E)} P_{N-1}^{(P)}) A^T Q_F C (F_{N-1}^{(E)})^* - ((F_{N-1}^{(E)})^*)^T (R_v \\ & - C^T Q_F C)(F_{N-1}^{(E)})^* P_{N-1}^{(E,P)} + \Gamma^T Q_F \Gamma Q_p\,) \end{aligned} \tag{8.50}$$

Similarly,

$$
\begin{aligned}
c_k^{(E)} &= c_{k+1}^{(E)} \\
&+ 2\, Trace((F_k^{(E)})^*)^T B^T (Q + P_{k+1} K_{k+1}^{(E)} H^{(E)}) A (P_k^{(E)} - \tilde{P}_k^{(E,P)}) \\
&+ Trace(A^T Q A P_k^{(E)}) + Trace(\Gamma^T Q \Gamma Q_p) \\
&+ Trace(\Gamma^T (H^{(E)})^T (K_{k+1}^{(E)})^T P_{k+1} K_{k+1}^{(E)} H^{(E)} \Gamma Q_p) \\
&+ Trace((K_{k+1}^{(E)})^T P_{k+1} K_{k+1}^{(E)} R_m^{(E)}) \\
&+ Trace(A^T (H^{(E)})^T (K_{k+1}^{(E)})^T P_{k+1} K_{k+1}^{(E)} H^{(E)} A P_k^{(E)}),\ k = N-2, \ldots, 0
\end{aligned}
\tag{8.51}
$$

and for $k = N-1$ we use the end-game equation

$$
\begin{aligned}
c_{N-1}^{(E)}(P_{N-1}^{(E)}; P_{N-1}^{(P)}, \tilde{P}_{N-1}^{(E,P)}) = Trace(\, & A^T Q_F A P_{N-1}^{(E)} + ((F_{N-1}^{(P)})^*)^T (R_u \\
&+ B^T Q_F B)(F_{N-1}^{(P)})^* P_{N-1}^{(E,P)} + 2((F_{N-1}^{(P)})^*)^T B^T Q_F A (P_{N-1}^{(E)} \\
&- \tilde{P}_{N-1}^{(E,P)} + \Gamma^T Q_F \Gamma Q_p)\,)
\end{aligned}
\tag{8.52}
$$

Remark. Only the parameters A, B, C, Q, Q_F, $R_c^{(P)}$ and $R_c^{(E)}$ feature in the Riccati equation for P_k, as if the game would be the deterministic LQDG. The players' measurement matrices, the process noise parameters and the measurement noise covariances do not feature in Eq. (8.57). However the solution P_k of the Riccati equation (8.57) and the LQGDG's measurements—related parameters $H^{(P)}$, $H^{(E)}$, the process noise parameters , $R_m^{(P)}$ and $R_m^{(E)}$, and the Kalman gains, all enter the recursions for the "intercepts" $c^{(P)}$ and $c^{(E)}$.

8.6 Main Result

The analysis of the LQGDG with a control-sharing information pattern is summarized in the following

Theorem 1. *Consider the LQGDG (8.1)–(8.6) with the information pattern:*

1. *The P and E players' prior information is given in Eqs. (8.2) and (8.4), respectively. The prior information $\overline{x}_0^{(P)}$ and $\overline{x}_0^{(E)}$ is private information of the respective P and E players and it is not shared among the P and E players. The covariances $P_0^{(P)}$ and $P_0^{(E)}$ are finite and are public information.*
2. *At decision time $1 \le k \le N-1$ the measurements of player P and player E are $z_k^{(P)}$ and $z_k^{(E)}$ and their measurement equations are Eqs. (8.3) and (8.5), respectively. At decision time $1 \le k \le N-1$ the respective measurement records*

$Z_k^{(P)} = \{z_1^{(P)}, \ldots, z_k^{(P)}\}$ and $Z_k^{(E)} = \{z_1^{(E)}, \ldots, z_k^{(E)}\}$ are the private information of players P and E and the measurements are not shared among the P and E players.

3. *At decision time $k = 1, \ldots, N-1$ the P and E players have complete recall of their respective ownship control histories $U_k = \{u_0, \ldots, u_{k-1}\}$ and $V_k = \{v_0, \ldots, v_{k-1}\}$.*
4. *The players observe their opponent's moves: at decision time $1 \le k \le N-1$ the control history $V_k = \{v_0, \ldots, v_{k-1}\}$ of player E is known to player P and, similarly, player E knows the control history $U_k = \{u_0, \ldots, u_{k-1}\}$ of player P.*

The players obtain their private state estimates $\bar{x}_k^{(P)}$ and $\bar{x}_k^{(E)}$ by running two *separate* Kalman Filters (KFs) in parallel driven by their private prior information and their separate measurements: Player P initialized his KF (8.11)–(8.15) with his prior information $(\bar{x}_0^{(P)}, P_0^{(P)})$ and uses his measurements $z_k^{(P)}$. Similarly, player E initialized his KF (8.16)–(8.20) with his prior information $(\bar{x}_0^{(E)}, P_0^{(E)})$ and uses his measurements $z_k^{(E)}$. Both players use the shared complete input history.

The players reuse the state feedback optimal strategies derived for the deterministic LQDG as provided by Theorem A1: In Eq. (8.61) player P sets $x_k := \bar{x}_k^{(P)}$ and in Eq. (8.62) player E sets $x_k := \bar{x}_k^{(E)}$.

A Nash equilibrium for the "zero-sum" LQGDG with a control-sharing information pattern is established. The value functions of players P and E are

$$V_k^{(P)}(\bar{x}_k^{(P)}, P_k^{(P)}; P_k^{(E)}, \tilde{P}_k^{(E,P)}) = (\bar{x}_k^{(P)})^T P_k \bar{x}_k^{(P)} + c_k^{(P)}$$
$$V_k^{(E)}(\bar{x}_k^{(E)}, P_k^{(E)}; P_k^{(P)}, \tilde{P}_k^{(E,P)}) = (\bar{x}_k^{(E)})^T P_k \bar{x}_k^{(E)} + c_k^{(E)}$$

where the matrices P_k are the solution of the Riccati equation (8.57). The "intercepts" $c_k^{(P)}$ and $c_k^{(E)}$ are obtained by solving the respective scalar recursions [(8.49), (8.50), (8.48)] and [(8.51), (8.52), (8.47)]. The covariance matrices $P_k^{(P)}$, $P_k^{(E)}$ and $\tilde{P}_k^{(E,P)}$ exclusively feature in the intercepts' recursions. The matrices $\tilde{P}_k^{(E,P)}$ are given by the solution of the Lyapunov-like linear matrix equation (8.29). The control Riccati equation (8.57), the KF Riccati equation (8.12), (8.13), (8.15) of player P, the KF Riccati equation (8.17), (8.18), (8.20) of player E , and the Lyapunov-like linear matrix equation (8.29) can all be solved ahead of time and off line. Once the three Riccati equations and the Lyapunov equation have been solved, the value functions' "intercepts" $c_k^{(P)}$ and $c_k^{(E)}$ are also obtained off line. □

8.7 Conclusion

Linear-Quadratic Gaussian Dynamic Games with a control-sharing information pattern have been considered. The players' initial state information and their measurements are private information, but each player is able to observe his antagonist's past inputs: the protagonists' past controls is shared information. Although this is a game with partial information, the control-sharing information pattern renders the game amenable to solution by the method of DP and a Nash equilibrium for the "zero-sum" LQGDG is established. The attendant optimal strategies of the LQGDG with a control-sharing information pattern are linear and certainty equivalence holds. The linearity of the optimal strategies has not been artificially imposed from the outset but follows from the LQG nature of the optimization problem at hand, courtesy of the control-sharing information pattern. The correct solution of LQGDGs with a control-sharing information pattern is obtained in closed-form.

Appendix: Linear-Quadratic Dynamic Game

The solution of Linear-Quadratic Dynamic Games (LQDG) with perfect information, a.k.a., deterministic LQDGs, was derived in Pachter and Pham (2010, Theorem 2.1). The Schur complement concept (Fuzhen 2005) was used to invert a blocked $(m_u + m_v) \times (m_u + m_v)$ matrix which contains four blocks, its two diagonal blocks being a $m_u \times m_u$ matrix and a $m_v \times m_v$ matrix. We further improve on the results of Pachter and Pham (2010) by noting that a matrix with four blocks has *two* Schur complements, say S_B and S_C. This allows one to obtain *explicit* and *symmetric* formulae for the P and E players' optimal strategies, thus yielding the *complete* solution of the deterministic LQDG. These results are used in this paper and for the sake of completeness, the closed form solution of the perfect information/deterministic zero-sum LQDG is included herein.

The linear dynamics are

$$x_{k+1} = Ax_k + Bu_k + Cv_k \ , \quad x_0 \equiv x_0 \ , \quad k = 0, 1, \ldots, N-1 \tag{8.53}$$

Payer P is the minimizer and his control $u_k \in R^{m_u}$. Player E is the maximizer and his control $v_k \in R^{m_v}$. The planning horizon is N. The cost/payoff functional is quadratic:

$$J(\{u_k\}_{k=0}^{N-1}, \{v_k\}_{k=0}^{N-1}; x_0) = x_N^T Q_F x_N + \sum_{k=0}^{N-1} (x_{k+1}^T Q x_{k+1} + u_k^T R_u u_k - v_k^T R_v v_k) \tag{8.54}$$

and Q and Q_F are real symmetric matrices. The players' control effort weighting matrices R_u and R_v are typically real symmetric and positive definite. Oftentimes it is stipulated that also the state penalty matrices Q and Q_F be positive definite, or, at least, positive semi-definite; these assumptions can be relaxed. The following holds.

Theorem A1. *A necessary and sufficient condition for the existence of a solution to the deterministic zero-sum LQDG (8.53) and (8.54) is*

$$R_u + B^T P_k B > 0 \tag{8.55}$$

and

$$R_v > C^T P_k C \tag{8.56}$$

$\forall\ k = 1, \ldots, N-1$, *where the real, symmetric matrices P_k are the solution of the Riccati difference equation*

$$\begin{aligned} P_{k+1} &= A^T\{P_k - P_k[BS_B^{-1}(P_k)B^T + BS_B^{-1}(P_k)B^T P_k C(R_v \\ &\quad - C^T P_k C)^{-1} C^T + C(R_v - C^T P_k C)^{-1} C^T P_k BS_B^{-1}(P_k)B^T \\ &\quad + C(R_v - C^T P_k C)^{-1} C^T P_k BS_B^{-1}(P_k)B^T P_k C(R_v \\ &\quad - C^T P_k C)^{-1} C^T + C(C^T P_k C - R_v)^{-1} C^T]P_k\}A + Q\,, \\ P_0 &= Q + Q_F\,,\;\; k = 0, \ldots, N-1 \end{aligned} \tag{8.57}$$

In Eq. (8.57), the first Schur complement matrix function

$$S_B(P_k) \equiv B^T P_k B + R_u + B^T P_k C(R_v - C^T P_k C)^{-1} C^T P_k B$$

In addition, the problem's parameters must satisfy the conditions

$$R_u + B^T(Q + Q_F)B > 0 \tag{8.58}$$

and

$$R_v > C^T(Q + Q_F)C \tag{8.59}$$

The value of the LQDG is

$$V_0(x_0) = x_0^T(P_N - Q)x_0 \tag{8.60}$$

The players' optimal strategies are the linear state feedback control laws

$$\begin{aligned} u_k^*(x_k) &= -S_B^{-1}(P_{N-k-1})B^T[I + P_{N-k-1}C(R_v \\ &\quad - C^T P_{N-k-1}C)^{-1}C^T]P_{N-k-1}A \cdot x_k, \end{aligned} \tag{8.61}$$

$$\begin{aligned} v_k^*(x_k) &= -S_C^{-1}(P_{N-k-1})C^T[I - P_{N-k-1}B(R_u \\ &\quad + B^T P_{N-k-1}B)^{-1}B^T]P_{N-k-1}A \cdot x_k \end{aligned} \tag{8.62}$$

In Eq. (8.62) the second Schur complement matrix function

$$\begin{aligned} S_C(P_{k+1}) &\equiv -\{R_v - C^T(Q + P_{k+1})C + C^T(Q \\ &\quad + P_{k+1})B[B^T(Q + P_{k+1})B + R_u]^{-1}B^T(Q + P_{k+1})C\} \end{aligned}$$

□

References

Aoki M (1973) On decentralized linear stochastic control. IEEE Trans Autom Control 18(3): 243–250

Basar T, Bernhard P (2008) H^{∞} - optimal control and related minimax design problems: a dynamic game approach. Birkhauser, Boston

Basar T, Olsder GJ (1995) Dynamic non-cooperative game theory. Academic, New York

Engwerda J (2005) LQ dynamic optimization and differential games. Wiley, New York

Fuzhen Z (2005) The Schur complement and its applications. Springer, Berlin

Ho YC, Bryson AE, Baron S (1965) Differential games and optimal pursuit-evasion strategies. IEEE Trans Autom Control 10(4):385–389

Kurtaran BZ, Sivan R (1974) Linear-quadratic-Gaussian control with one-step-delay sharing pattern. IEEE Trans Autom Control 13(5):571–574

Pachter M (2013) Static linear-quadratic Gaussian games. In: Krivan V, Zaccour G (eds) Advances in dynamic games. Birkhauser, Boston, pp 85–105

Pachter M, Pham K (2010) Discrete-time linear-quadratic dynamic games. J Optim Theory Appl 146(1):151–179

Pachter M, Pham K (2013) Static teams and stochastic games. In: Pardalos PM, et al (eds) Dynamics of information systems: algorithmic approaches. Springer, New York, pp 147–176

Pachter M, Pham K (2014) Informational issues in decentralized control. In: Vogiatzis C, et al (eds) Dynamics of information systems: computational and mathematical challenges. Springer, New York, pp 45–76

Radner R (1962) Static teams and stochastic games. Ann Math Stat 33(3):857–867

Sandell NR, Athans M (1974) Solution of some nonclassical LQG stochastic decision problems. IEEE Trans Autom Control 19(2):108–116

Witsenhausen HS (1968) A counterexample in stochastic optimal control. SIAM J Control 6(1):131–147

Witsenhausen HS (1971) Separation of estimation and control for discrete time systems. Proc IEEE 59(11):1557–1566

Chapter 9
Pursuit-Evasion Game of Kind Between Hybrid Players

Josef Shinar, Valery Y. Glizer, and Vladimir Turetsky

Abstract A pursuit-evasion differential game of kind with bounded controls and prescribed duration is considered. Both players have two possible dynamics and both can switch between them once during the game. Each player knows the two possible dynamics of the other, but not the actual one. The optimal strategies of the players in this game include the order of the two modes and the time of the mode change between them. The optimal use of the mode change enlarges the winning zone of the respective player, compared to its winning zone when using fixed dynamics. An algorithmic example illustrates the complexity of the game with hybrid players.

Keywords Pursuit-evasion game • Hybrid dynamics • Capture conditions • Capture zone

MSC Classification Codes: 49N70, 91A23, 93C30

9.1 Introduction

The optimal performance of interceptor missiles against maneuverable targets can be analyzed by using the mathematical model of pursuit-evasion differential games. The complete solutions of planar linear pursuit-evasion games with bounded

J. Shinar
Faculty of Aerospace Engineering, Technion - Israel Institute of Technology, Technion City, Haifa 3200003, Israel
e-mail: shinarj@technion.ac.il

V.Y. Glizer • V. Turetsky (✉)
Department of Applied Mathematics, Ort Braude College of Engineering, 51 Snunit Str., P.O.B. 78, Karmiel 2161002, Israel
e-mail: valery48@braude.ac.il; turetsky1@braude.ac.il

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_9

controls and linear-quadratic pursuit-evasion games are well known (Ho et al. 1965; Gutman and Leitmann 1976; Gutman 1979; Shinar 1981; Shima and Shinar 2002; Turetsky and Shinar 2003; Shinar et al. 2013).

Modern flying vehicles can use two modes to control their trajectories. The aerodynamic mode, depending on the vehicle's speed and altitude, can be used only in the lower atmosphere. The magnitude of the lateral acceleration, created by the respective lift force, can be rather large, but its dynamics, depending on the aerodynamic configuration, can be rather slow. In high altitudes, where the air density is low, as well as out of the atmosphere, only thrust vector control (TVC) can be used. The magnitude of the lateral acceleration, created by TVC is rater limited, but its dynamics is rather fast. Of course in the lower altitudes both control modes can be applied. Changing the dynamics is an additional element in the hybrid flying vehicle's control options.

During recent decades, control problems of hybrid dynamics systems were studied by many researchers. In the works devoted to this topic, mostly control problems with a single decision maker are studied (see Utkin 1983; Bartolini and Zolezzi 1986; Chen and Fukuda 1997; Sussmann 1999; Lee and Kouvaritakis 2000; Riedinger et al. 2003; Choi 2004 and the references therein). Control problems with two and more decision makers (games with hybrid dynamics) are investigated much less. In Grigorenko (1991) a differential game of pursuit of a single evader by a group of pursuers is considered. The structure of the game dynamics is changed by the evader once during the game. Sufficient conditions for the existence of the game solution are obtained. In Mitchell et al. (2001, 2005) the reachability sets for pursuit-evasion games with nonlinear hybrid dynamics are numerically constructed by using solutions of time-dependent Hamilton-Jacobi equations. In Gao et al. (2007) a general pursuit-evasion differential game with hybrid dynamics is studied using the viability theory and non-smooth analysis.

The pursuit-evasion game, considered in the present paper, is the mathematical model of an interception engagement between two vehicles, an interceptor P (*pursuer*) and its target E (*evader*) both moving with constant velocities in a horizontal plane. The dynamics of each vehicle is approximated by a first-order linear transfer function with time constants τ_p and τ_e, respectively. Moreover, it is assumed that the players' controls (their lateral acceleration commands) are bounded by the constants $a_p^{\max}$ and $a_e^{\max}$, respectively. Thus, the dynamics of each player is defined by the respective vector $\omega_i = (a_i^{\max}, \tau_i)$, $i = p, e$. The cost function of the corresponding game of degree is the distance of closest approach (miss distance). The fixed dynamics version of this game and its extension to the case of time-varying velocities and lateral acceleration command bounds have been studied extensively in the open literature, see Shinar (1981), Shima and Shinar (2002), Gutman (2006), and Glizer and Turetsky (2008). It was shown that this game has a saddle-point solution in feedback strategies. The solution leads to the decomposition of the game space into two regions (singular and regular) of different optimal strategies. The game space decomposition is completely determined by the pair (ω_p, ω_e). This pair also determines the existence or non-existence of a capture zone—the set of all initial positions of the game for which the value of the game of degree equals zero. In other words this pair determines the playability of the game of kind.

In Shinar et al. (2009b), the pursuit of an evader with fixed dynamics by a pursuer with hybrid dynamics is studied, while in Shinar et al. (2009a), the evasion from a pursuer with fixed dynamics by an evader with hybrid dynamics is analyzed. In these papers, it was established that using the hybrid dynamics is helpful only if the capture zones' boundaries of the possible fixed dynamics games intersect. The conditions of the intersection, as well as the optimal order of dynamics for the player with hybrid dynamics were also determined. Moreover, in each case, the optimal time of the mode change was derived in a closed form. It is important to note that this optimal time depends only on the dynamic modes of the hybrid player. The capture zone of the hybrid pursuer and the escape zone of the hybrid evader also were constructed.

In the present paper, the case of a hybrid pursuer and a hybrid evader is treated from the viewpoint of the pursuer. The players can change their dynamics from one mode to another once during the game. The order of the dynamics and the time of the change (the dynamic schedule) are additional elements of player's control, unknown to the opponent. The hybrid dynamics, unknown to the respective opponent, makes this game a differential game with incomplete information (Shinar et al. 2009c, 2010, 2012). Such games were considered, for instance, in Krasovskii (1984), Petrosjan (1993), Kumkov and Patsko (1995), and Chernousko and Melikyan (1975).

The chapter is organized as follows. The next section is devoted to the problem formulation. In Sect. 9.3, main results of the fixed dynamics games are briefly summarized. Section 9.4 contains the solution of the pursuit problem of a hybrid evader by a fixed dynamics pursuer. For a hybrid pursuer two such phases can be considered, before and after the change of its dynamic mode. Section 9.5 presents the two phases assuming fixed dynamic schedules, leading to determine the game optimal (minmax) schedules of each player, and it presents an example illustrating the algorithmic solution methodology. Conclusions are summarized in Sect. 9.6. The proofs of technical lemmas are presented in the Appendices.

9.2 Problem Statement

9.2.1 Engagement Model

A planar engagement between two moving objects (players)—a pursuer and an evader—is considered. The schematic view of this engagement is shown in Fig. 9.1. The X axis of the coordinate system is aligned with the initial line of sight. The origin is collocated with the initial pursuer position. The points (x_p, y_p), (x_e, y_e) are the current coordinates, V_p and V_e are the velocities and a_p, a_e are the lateral accelerations of the pursuer and the evader respectively, φ_p, φ_e are the respective

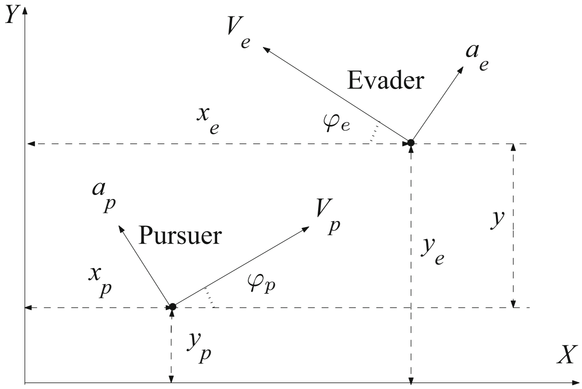

Fig. 9.1 Interception geometry

angles between the velocity vectors and the reference line of sight; and $y = y_e - y_p$ is the separation normal to the initial line of sight.

It is assumed that the dynamics of each object is expressed by a first-order transfer function with the time constants τ_p and τ_e, respectively. The velocities and the bounds of the lateral acceleration commands of both objects are constant.

If the aspect angles φ_p and φ_e are small during the engagement then the linearized engagement model is (Shinar 1981)

$$\dot{x} = Ax + bu + cv, \quad x(0) = x_0, \tag{9.1}$$

where the state vector is $x = (x_1, x_2, x_3, x_4)^T = (y, \dot{y}, a_e, a_p)^T$, the superscript T denotes the transposition,

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1/\tau_e & 0 \\ 0 & 0 & 0 & -1/\tau_p \end{bmatrix}, \tag{9.2}$$

$$b = (0, 0, 0, a_p^{\max}/\tau_p)^T, \quad c = (0, 0, a_e^{\max}/\tau_e, 0)^T, \tag{9.3}$$

$$x_0 = (0, x_{20}, 0, 0)^T, \;\; x_{20} = V_e\varphi_e(0) - V_p\varphi_p(0). \tag{9.4}$$

The small angles assumption allows one to calculate the final time of the engagement as

$$t_f = r_0/(V_p + V_e), \tag{9.5}$$

where r_0 is the initial range between the players.

The normalized lateral acceleration commands of the evader $v(t)$ and the pursuer $u(t)$ satisfy the constraints

$$|v(t)| \leq 1, \quad |u(t)| \leq 1, \quad 0 \leq t \leq t_f. \tag{9.6}$$

Remark 1. The engagement model (9.1)–(9.3) is completely determined by the two vectors $\omega_e = (a_e^{\max}, \tau_e)$ and $\omega_p = (a_p^{\max}, \tau_p)$ called in the sequel the *dynamic modes* of the evader and pursuer, respectively.

9.2.2 Pursuit-Evasion Game Between Hybrid Players

It is assumed that both the pursuer and the evader have two dynamic modes at their disposal. Thus, in (9.1)–(9.3), $\omega_e \in \Omega_e \triangleq \{\omega_e^1,\ \omega_e^2\}$, $\omega_p \in \Omega_p \triangleq \{\omega_p^1,\ \omega_p^2\}$, where $\omega_e^j = (a_{ej}^{\max}, \tau_{ej})$, $j = 1, 2$, $\omega_p^i = (a_{ei}^{\max}, \tau_{ei})$, $i = 1, 2$ are given dynamic modes of the evader and the pursuer. The pursuer and the evader can switch dynamics from one mode to another once during the engagement. This means that they choose the numbers $j_1, j_2 \in \{1, 2\}$, $i_1, i_2 \in \{1, 2\}$, and the time moments $t_e, t_p \in [0, t_f]$, such that

$$\omega_e = \begin{cases} \omega_e^{j_1}, \ 0 \leq t \leq t_e, \\ \\ \omega_e^{j_2}, \ t_e < t \leq t_f, \end{cases} \qquad \omega_p = \begin{cases} \omega_p^{i_1}, \ 0 \leq t \leq t_p, \\ \\ \omega_p^{i_2}, \ t_p < t \leq t_f. \end{cases} \tag{9.7}$$

The triplets $s_e \triangleq (j_1, j_2, t_e)$ and $s_p \triangleq (i_1, i_2, t_p)$ are called the *dynamic schedule* of the evader and the pursuer, respectively.

In the game of degree the cost function is the miss distance

$$J = |x_1(t_f)|. \tag{9.8}$$

The objective of the pursuer is to minimize the miss distance by means of the feedback strategy $u(t, x)$ and the dynamic schedule s_p against optimal (worst case) evader strategy $v(t, x)$ and dynamic schedule s_e. It is assumed that the pursuer knows the evader set Ω_e and the current state of the engagement $x(t)$, but not the current dynamic mode of the evader. This is the pursuer's view of the Hybrid Pursuit Game (HPG).

Remark 2. If, in particular, $t_e = 0$ or $t_e = t_f$, the dynamics of the evader is constant during the engagement. However, the pursuer is not aware, which dynamics of two possible the evader uses. This constant evader dynamics pursuit-evasion problem with incomplete information forms a special case of the HPG.

9.3 Fixed Dynamics Pursuit-Evasion Problem

In this section, the case, where the dynamic modes of the players are fixed, is briefly presented. In the sequel, this problem is called the Original Fixed Dynamics Game (OFDG). This game was solved in Shinar (1981). Its solution is stated in the sequel.

9.3.1 Zero-Effort Miss Distance

The solution of the OFDG is based on its scalarization by introducing a new state variable

$$Z(t) = Z(t; \omega_e, \omega_p) = d^T \Phi(t_f, t; \tau_e, \tau_p) x(t; \omega_e, \omega_p), \tag{9.9}$$

where $x(t; \omega_e, \omega_p)$ is the state vector of (9.1), $\Phi(t_f, t; \tau_e, \tau_p)$ is the transition matrix of the homogeneous system $\dot{x} = Ax$ and $d^T = (1, 0, 0, 0)$. The value of the function $Z(t)$ has the following physical interpretation. If $u \equiv 0$ and $v \equiv 0$ on the interval $[t, t_f]$, then the miss distance $|x_1(t_f)|$ equals $|Z(t)|$. Therefore, this function is called the zero-effort miss distance (ZEM). It is given explicitly by

$$\begin{aligned} Z(t) = x_1(t) + (t_f - t)x_2(t) + \\ \tau_e^2 \Psi\left((t_f - t)/\tau_e\right) x_3(t) - \tau_p^2 \Psi\left((t_f - t)/\tau_p\right) x_4(t), \end{aligned} \tag{9.10}$$

where $\Psi(\xi) \triangleq \exp(-\xi) + \xi - 1 > 0$, $\xi > 0$. By introducing a new independent variable $\vartheta = t_f - t$ (time-to-go) and using (9.10), it can be shown that the function of ϑ

$$\tilde{Z}(\vartheta) = \tilde{Z}(\vartheta; \omega_e, \omega_p) \triangleq Z(t_f - \vartheta; \omega_e, \omega_p) \tag{9.11}$$

satisfies the differential equation

$$\frac{d\tilde{Z}}{d\vartheta} = h(\vartheta, \tau_p, a_p^{\max}) u(t_f - \vartheta) - h(\vartheta, \tau_e, a_e^{\max}) v(t_f - \vartheta), \tag{9.12}$$

where

$$h(\vartheta, \tau, a^{\max}) = \tau a^{\max} \Psi(\vartheta/\tau). \tag{9.13}$$

Note that $\tilde{Z}(0) = x_1(t_f)$, i.e., the performance index (9.8) can be rewritten as $J = |\tilde{Z}(0)|$. This allows to associate the OFDG with a scalar game consisting of the dynamics

$$\frac{dz}{d\vartheta} = h(\vartheta, \tau_p, a_p^{\max})u - h(\vartheta, \tau_e, a_e^{\max})v, \;\; z(\vartheta_0) = z_0, \tag{9.14}$$

the performance index

$$J = |z(0)|, \tag{9.15}$$

the control constraints

$$|v(\vartheta)| \leq 1, \;\; |u(\vartheta)| \leq 1, \;\; 0 \leq \vartheta \leq \vartheta_0, \tag{9.16}$$

where

$$\vartheta_0 = t_f, \quad z_0 = \vartheta_0 x_{20}, \tag{9.17}$$

the pursuer and evader controls $u(\vartheta)$ and $v(\vartheta)$ are actually $u(t_f - \vartheta)$ and $v(t_f - \vartheta)$ of the OFDG.

The game described by (9.14)–(9.16) is called the Scalar Fixed Dynamics Game (SFDG).

9.3.2 SFDG Solution

The solution of this problem is based on the decomposition of the game space $D \triangleq \{(\vartheta, z) : \vartheta \in [0, \vartheta_0], \; z \in R^1\}$ into two regions of different strategies.

In the first (*regular*) region D_1 the optimal pursuer strategy and the worst case evader strategy have the "bang-bang" structure:

$$u^0(\vartheta, z) = v^0(\vartheta, z) = \text{sign}(z), \tag{9.18}$$

and the guaranteed result is nonzero, depending on the initial conditions. In the second (*singular*) region $D_0 = D \backslash D_1$ the optimal pursuer strategy $u^0(\vartheta, z)$ and the worst case evader strategy $v^0(\vartheta, z)$ are *arbitrary* subject to (9.16) and the guaranteed pursuit result is constant. Note that D_1 and D_0 are symmetrical with respect to the ϑ-axis.

If the following two inequalities, called the "capture conditions", are satisfied the pursuer can achieve zero miss distance (*capture*), from a part of the game space (*capture zone*) and the corresponding game of kind is *playable*.

In this case

$$a_p^{\max} > a_e^{\max}, \;\; a_p^{\max}/\tau_p \geq a_e^{\max}/\tau_e, \tag{9.19}$$

and the singular region is

$$D_0 = D_0(\omega_e, \omega_p) = \{(\vartheta, z) \in D : \ |z| < z^*(\vartheta, \omega_e, \omega_p)\}, \tag{9.20}$$

where

$$z^*(\vartheta, \omega_e, \omega_p) = \int_0^{\vartheta} H(\xi, \omega_e, \omega_p) d\xi, \tag{9.21}$$

$$H(\xi, \omega_e, \omega_p) \triangleq h(\xi, \tau_p, a_p^{\max}) - h(\xi, \tau_e, a_e^{\max}). \tag{9.22}$$

In this case, for any initial position (ϑ_0, z_0) the guaranteed result is given by

$$J^0 = J^0(\vartheta_0, z_0, \omega_e, \omega_p) =$$
$$\begin{cases} 0, & (\vartheta_0, z_0) \in D_0(\omega_e, \omega_p), \\ |z_0| - \int_0^{\vartheta_0} H(\vartheta, \omega_e, \omega_p) d\vartheta, & (\vartheta_0, z_0) \in D_1(\omega_e, \omega_p). \end{cases} \tag{9.23}$$

Thus, subject to the conditions (9.19), the closure of the singular region clo(D_0) becomes the *robust capture zone*, i.e. the set of all initial positions, from which the pursuer can guarantee zero miss distance against any admissible evader strategy. In Fig. 9.2, the singular and the regular regions are shown for the parameters $\tau_p = 0.2$ s, $a_p^{\max} = 180$ m/s^2, $\tau_e = 0.4$ s, $a_e^{\max} = 130$ m/s^2, $t_f = 5$ s.

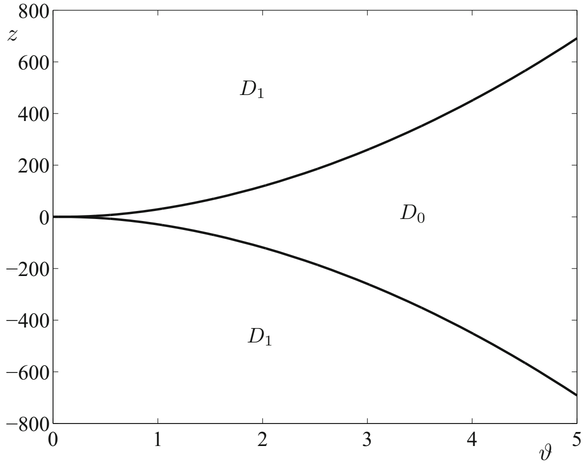

Fig. 9.2 The SFDG space decomposition

9.4 Transformation of the HPG

In the sequel, it is assumed that the "capture conditions" (9.19) are satisfied for each pair (ω_e^j, ω_p^i), $j, i = 1, 2$, i.e.

$$a_{pi}^{\max} > a_{ej}^{\max}, \quad a_{pi}^{\max}/\tau_{pi} \geq a_{ej}^{\max}/\tau_{ej}, \quad j, i = 1, 2. \tag{9.24}$$

9.4.1 New Aiming Point of the Pursuer

Due to (9.10)–(9.11),

$$\tilde{Z}(\vartheta) = \tilde{Z}_0(\vartheta) + \tau_e^2 \Psi(\vartheta/\tau_e)\tilde{x}_3(\vartheta), \tag{9.25}$$

where

$$\tilde{Z}_0(\vartheta) = \tilde{x}_1(\vartheta) + \vartheta \tilde{x}_2(\vartheta) - \tau_p^2 \Psi(\vartheta/\tau_p)\tilde{x}_4(\vartheta), \tag{9.26}$$

$$\tilde{x}_i(\vartheta) = x_i(t_f - \vartheta), \quad i = 1, \ldots, 4. \tag{9.27}$$

Note that in the OFDG, the pursuer knows the current ZEM value perfectly, and it serves as an aiming point in this game. In the HPG, the pursuer has information only on the two possible current values of the ZEM, but does not know the actual one. Thus, $\tilde{Z}(\vartheta)$ cannot serve as an aiming point in the HPG as in the OFDG. A new aiming point is defined based on the concept of an uncertainty set (Kumkov and Patsko 1995; Petrosjan 1993). For any $\vartheta \in [0, \vartheta_0]$, the ZEM uncertainty set consists of the two possible current values of the ZEM: $\tilde{\mathcal{Z}}(\vartheta) \triangleq \{\tilde{Z}(\vartheta, \omega_e^1, \omega_p), \tilde{Z}(\vartheta, \omega_e^2, \omega_p)\}$, where, due to (9.25),

$$\tilde{Z}(\vartheta, \omega_e^j, \omega_p) = \tilde{Z}_0(\vartheta) + \tau_{ej}^2 \Psi(\vartheta/\tau_{ej})\tilde{x}_3(\vartheta), \quad j = 1, 2. \tag{9.28}$$

The convex hull $\mathcal{C}_Z(\vartheta) = \mathrm{conv}(\tilde{\mathcal{Z}}(\vartheta))$ is the closed interval with the end points $\tilde{Z}(\vartheta, \omega_e^1, \omega_p), \tilde{Z}(\vartheta, \omega_e^2, \omega_p)$. Therefore, the center $z_c(\vartheta)$ of $\mathcal{C}_Z$ is given by

$$z_c(\vartheta) = \tilde{Z}_0(\vartheta) + F(\vartheta)\tilde{x}_3(\vartheta), \tag{9.29}$$

where

$$F(\vartheta) \triangleq \Big(\tau_{e1}^2 \Psi(\vartheta/\tau_{e1}) + \tau_{e2}^2 \Psi(\vartheta/\tau_{e2})\Big)\Big/2. \tag{9.30}$$

We choose $z_c(\vartheta)$ as the pursuer's aiming point in the HPG.

Remark 3. For any $\vartheta \in [0, \vartheta_0]$, let us consider an auxiliary game, in which the players choose the points $Z_p, Z_e \in \mathcal{C}_Z(\vartheta)$ according to the performance index

$$|Z_p - Z_e| \to \min_{Z_p \in \mathcal{C}_Z(\vartheta)} \max_{Z_e \in \mathcal{C}_Z(\vartheta)} . \tag{9.31}$$

This game is a particular case of the one considered in Petrosjan (1993). The value of $z_c(\vartheta)$ is the pursuer's optimal choice in this game. The optimal evader's choice in this game is mixed: choosing one of the ends of $\mathcal{C}_Z$ with equal probabilities $1/2$. Thus, $z_c(\vartheta)$ is the guaranteed minimal distance between the new aiming point of the pursuer and the actual ZEM value.

Note that

$$z_c(0) = x_1(t_f). \tag{9.32}$$

By direct differentiation, the new aiming point, along with $\tilde{x}_3$, satisfies the system of differential equations

$$\frac{dz_c}{d\vartheta} = h(\vartheta, \tau_p, a_p^{\max})u - h_e(\vartheta, \tau_e, a_e^{\max})v + h_x(\vartheta, \tau_e)\tilde{x}_3, \tag{9.33}$$

$$\frac{d\tilde{x}_3}{d\vartheta} = \frac{1}{\tau_e}(\tilde{x}_3 - a_e^{\max} v), \tag{9.34}$$

where $h(\vartheta, \tau_p, a_p^{\max})$ is given by (9.13), and

$$h_e(\vartheta, \tau_e, a_e^{\max}) = F(\vartheta)\frac{a_e^{\max}}{\tau_e}, \tag{9.35}$$

$$h_x(\vartheta, \tau_e) = \frac{F(\vartheta)}{\tau_e} - G(\vartheta), \tag{9.36}$$

$$G(\vartheta) \triangleq \Big(\tau_{e1}\Psi(\vartheta/\tau_{e1}) + \tau_{e2}\Psi(\vartheta/\tau_{e2})\Big)\Big/2. \tag{9.37}$$

Remember that in (9.34), $\omega_e = \omega_e^{j_1}$ for $\vartheta \in [\vartheta_e, \vartheta_0]$, while $\omega_e = \omega_e^{j_2}$ for $\vartheta \in [0, \vartheta_e)$.

Due to (9.4), (9.26), (9.27) and (9.29), the initial conditions for (9.33)–(9.34) are

$$z_c(\vartheta_0) = z_0, \quad \tilde{x}_3(\vartheta_0) = 0. \tag{9.38}$$

Based on the new aiming point, a new pursuit game is formulated for the system (9.33)–(9.38), the control constraints (9.16) and the performance index

$$\mathcal{J} = |z_c(0)| \to \min_u \max_v . \tag{9.39}$$

This game is also with hybrid dynamics. However, in contrast with the ZEM in the HPG, the new state variable is known perfectly by the pursuer. This new game is called the Reduced Hybrid Pursuit Game (RHPG).

9.4.2 Impulsive Dynamics Pursuit Game (IPG)

By virtue of (9.26) and (9.29), the new state variable $z_c(\vartheta)$ has a jump at $\vartheta = \vartheta_p$

$$\Delta z_c\Big|_{\vartheta=\vartheta_p} = \delta\tilde{x}_4(\vartheta_p), \tag{9.40}$$

where

$$\delta = \tau_{p1}^2\Psi(\vartheta_p/\tau_{p1}) - \tau_{p2}^2\Psi(\vartheta_p/\tau_{p2}). \tag{9.41}$$

Since the jump (9.40) depends on $\tilde{x}_4(\vartheta)$, in the new game formulation the dynamics (9.33)–(9.34) should be augmented by the differential equation for $\tilde{x}_4(\vartheta)$. Thus, by using (9.1), the new game dynamics is described by the system

$$\begin{aligned}
\frac{dz_c}{d\vartheta} &= h(\vartheta, \tau_p, a_p^{\max})u - h_e(\vartheta, \tau_e, a_e^{\max})v + h_x(\vartheta, \tau_e)\tilde{x}_3, \; z_c(\vartheta_0) = z_0 \\
\frac{d\tilde{x}_3}{d\vartheta} &= \frac{1}{\tau_e}(\tilde{x}_3 - a_e^{\max}v) \qquad \tilde{x}_3(\vartheta_0) = 0, \\
\frac{d\tilde{x}_4}{d\vartheta} &= \frac{1}{\tau_p}(\tilde{x}_4 - a_p^{\max}u) \qquad \tilde{x}_4(\vartheta_0) = 0.
\end{aligned} \tag{9.42}$$

Let $\mathcal{S}_p$ be the set of all triplets $s_p = \{i_1, i_2, \vartheta_p\}$, $i_1, i_2 \in \{1, 2\}$, $i_1 \neq i_2$, $\vartheta_p \in [0, \vartheta_0]$. The triplet $s_p \in \mathcal{S}_p$ is called the *pursuer dynamics schedule*. Similarly, $\mathcal{S}_e$ is the set of all triplets $s_e = \{j_1, j_2, \vartheta_e\}$, $j_1, j_2 \in \{1, 2\}$, $j_1 \neq j_2$, $\vartheta_e \in [0, \vartheta_0]$, and the triplet $s_e \in \mathcal{S}_e$ is called the *evader dynamics schedule*.

The pursuer choice of $s_p \in \mathcal{S}_p$ and the evader choice of $s_e \in \mathcal{S}_p$ mean that

$$\omega_e = \begin{cases} \omega_e^{j_1}, \; \vartheta_0 \geq \vartheta > \vartheta_e \\ \omega_e^{j_2}, \; \vartheta_e \geq \vartheta \geq 0, \end{cases} \qquad \omega_p = \begin{cases} \omega_p^{i_1}, \; \vartheta_0 \geq \vartheta > \vartheta_p \\ \omega_p^{i_2}, \; \vartheta_p \geq \vartheta \geq 0, \end{cases} \tag{9.43}$$

Following (9.40), the z_c-jump can be represented as

$$\Delta z_c = \Delta z_c(s_p) = \delta\tilde{x}_4(\vartheta_p), \tag{9.44}$$

where

$$\delta = \tau_{p,i_1}^2 \Psi(\vartheta_p/\tau_{p,i_1}) - \tau_{p,i_2}^2 \Psi(\vartheta_p/\tau_{p,i_2}). \tag{9.45}$$

Therefore, for a given pursuer dynamics schedule s_p, the new dynamics (9.42) is subject to a jump

$$z_c(\vartheta_p) = z_c(\vartheta_p + 0) + \Delta z_c(s_p). \tag{9.46}$$

Thus, on the intervals $(\vartheta_p, \vartheta_0]$ and $[0, \vartheta_p)$, the new dynamics is described by the system (9.42), while at the point $\vartheta = \vartheta_p$, this dynamics is described by the jump condition (9.46), i.e. the new dynamics has an *impulsive* character. The cost function and control constraints in the new pursuit game remain the same as in the RHPG, i.e. (9.39) and (9.16), respectively. This game is called the Impulsive dynamics Pursuit Game (IPG).

9.5 IPG Solution

9.5.1 Optimal Pursuer Strategy on $[\vartheta_p, \vartheta_0]$ *for Fixed* s_p, s_e

For fixed schedules s_p, s_e, the optimal pursuer strategy on the interval $[\vartheta_p, \vartheta_0]$ is constructed based on the Auxiliary Game-1 (AG1) with the dynamics (9.42) for $\tau_p = \tau_{p,i_1}$, $a_p^{\max} = a_{p,i_1}^{\max}$, the control constraints (9.16) and the performance index

$$J_{\vartheta_p} \triangleq |z_c(\vartheta_p) + \delta \tilde{x}_4(\vartheta_p)| \rightarrow \min_u \max_v . \tag{9.47}$$

Note that J_{ϑ_p} is the absolute value of z_c for $\vartheta = \vartheta_p$ (after the jump). Due to Glizer and Turetsky (2008), the AG1 is equivalent to the Scalar Auxiliary Game-1 (SAG1) with the state variable

$$w(\vartheta) = w(\vartheta, z_c(\vartheta), \tilde{x}_3(\vartheta), \tilde{x}_4(\vartheta)) = d^T \Lambda(\vartheta_p, \vartheta) \begin{bmatrix} z_c(\vartheta) \\ \tilde{x}_3(\vartheta) \\ \tilde{x}_4(\vartheta) \end{bmatrix}, \tag{9.48}$$

where

$$d^T = [1, 0, \ \delta], \tag{9.49}$$

and $\Lambda(\vartheta, \sigma)$ is the fundamental matrix of the homogenous system, corresponding to (9.42), for $\tau_p = \tau_{p,i_1}$:

$$\Lambda(\vartheta,\sigma)=\begin{bmatrix} 1 \int\limits_{\sigma}^{\vartheta} h_x(\xi,\tau_e)\exp((\xi-\sigma)/\tau_e)d\xi & 0 \\ 0 \quad \exp((\vartheta-\sigma)/\tau_e) & 0 \\ 0 \quad 0 & \exp((\vartheta-\sigma)/\tau_{p,i_1}) \end{bmatrix}. \tag{9.50}$$

By virtue of (9.48),

$$J_{\vartheta_p}=|w(\vartheta_p)|. \tag{9.51}$$

Due to (9.48) and (9.50),

$$w(\vartheta)=z_c(\vartheta)+\tilde{x}_3(\vartheta)\int\limits_{\vartheta}^{\vartheta_p} h_x(\xi,\tau_e)\exp((\xi-\vartheta)/\tau_e)d\xi+ \\ \delta\exp\left((\vartheta_p-\vartheta)/\tau_{p,i1}\right)\tilde{x}_4(\vartheta). \tag{9.52}$$

By using (9.42) and (9.52), $w(\vartheta)$ satisfies the differential equation

$$\frac{dw}{d\vartheta}=h_{p1}(\vartheta,\vartheta_p,\tau_{p,i_1},\tau_{p,i_2},a_{p,i_1}^{\max})u-h_{e1}(\vartheta,\vartheta_p,\tau_e,a_e^{\max})v, \tag{9.53}$$

where

$$h_{p1}(\vartheta,\vartheta_p,\tau_{p,i_1},\tau_{p,i_2},a_{p,i_1}^{\max})=h(\vartheta,\tau_{p,i_1},a_{p,i_1}^{\max})-\frac{\delta a_{p,i_1}^{\max}}{\tau_{p,i_1}}\exp((\vartheta_p-\vartheta)/\tau_{p,i_1}), \tag{9.54}$$

$$h_{e1}(\vartheta,\vartheta_p,\tau_e,a_e^{\max})=h_e(\vartheta,\tau_e,a_e^{\max})+\frac{a_e^{\max}}{\tau_e}\int\limits_{\vartheta}^{\vartheta_p} h_x(\xi,\tau_e)\exp((\xi-\vartheta)/\tau_e)d\xi. \tag{9.55}$$

The initial condition is

$$w(\vartheta_0)=z_0. \tag{9.56}$$

Thus, the SAG1 has the dynamics (9.53) with the initial condition (9.56), the control constraints (9.16) and the cost function (9.51), to be minimized by the pursuer and maximized by the evader.

Remark 4. If $\vartheta_e>\vartheta_p$, the coefficient function $(-\hat{h}_{e1}(\vartheta,\vartheta_p,\tau_e,a_e^{\max}))$ for v in (9.53) is piecewise continuous on $(\vartheta_p,\vartheta_0]$. Otherwise, it is continuous.

By virtue of Glizer and Turetsky (2008), the SAG1 solution is based on the decomposition of the game space (ϑ,w) into two regions of different feedback controls. In the first (*singular*) region $\hat{D}_0$ the optimal feedback controls $\hat{u}^*(\vartheta,w)$ and

$\hat{v}^*(\vartheta, w)$ are *arbitrary* subject to (9.16). In the second (*regular*) region $\hat{D}_1 = R^2 \backslash \hat{D}_0$ the optimal feedback controls have a "bang-bang" structure:

$$\hat{u}^*(\vartheta, w) = \hat{v}^*(\vartheta, w) = \operatorname{sign} w. \tag{9.57}$$

Since $w(\vartheta_0) = z_0$, the value of SAG1 is given by

$$\hat{J}^* = \hat{J}^*(\vartheta_0, z_0, s_p) = \begin{cases} C_1 = \text{const}, & (\vartheta_0, z_0) \in \hat{D}_0, \\ |z_0| + \int_{\vartheta_0}^{\vartheta_p} H_1(\xi, s_p) d\xi, & (\vartheta_0, z_0) \notin \hat{D}_0, \end{cases} \tag{9.58}$$

where

$$H_1(\vartheta, s_p) \triangleq h_{p1}(\vartheta, \vartheta_1, \tau_{p,i_1}, \tau_{p,i_2}, a_{p,i_1}^{\max}) - h_{e1}(\vartheta, \vartheta_p, \tau_e, a_e^{\max}). \tag{9.59}$$

Remark 5. For the sake of simplicity, we use in the further analysis the bang-bang strategies (9.57) in the entire SAG1 space.

9.5.2 *Optimal Pursuer Strategy on* $[0, \vartheta_p]$ *for Fixed* s_p, s_e

The optimal pursuer strategy on the interval $[0, \vartheta_p]$ is constructed based on the Auxiliary Game-2 (AG2) with the dynamics (9.33)–(9.34) for $\tau_p = \tau_{p,i_2}$, $a_p^{\max} = a_{p,i_2}^{\max}$, the control constraints (9.16) and the performance index (9.39). The initial conditions are

$$z_c\Big|_{\text{AG2}}(\vartheta_p) = w(\vartheta_p), \;\; \tilde{x}_3\Big|_{\text{AG2}}(\vartheta_p) = \tilde{x}_3\Big|_{\text{AG1}}(\vartheta_p) \triangleq \tilde{x}_{30}. \tag{9.60}$$

Remark 6. Like in (9.53), if $\vartheta_e < \vartheta_p$, the coefficient functions in Eq. (9.33) $\big(-h_e(\vartheta, \tau_e, a_e^{\max})$ for v and $h_x(\vartheta, \tau_e)$ for $\tilde{x}_3\big)$ are piecewise continuous on $[0, \vartheta_p)$. Otherwise, they are continuous.

The game AG2 is solved based on the results of Shinar et al. (2012). Namely, let define two integers $m, M \in \{1, 2\}$ as follows:

$$\tau_{em} = \min\{\tau_{e1}, \tau_{e2}\}, \quad \tau_{eM} = \max\{\tau_{e1}, \tau_{e2}\}. \tag{9.61}$$

Let the following conditions hold:

$$a_{p,i_2}^{\max}/a_{em}^{\max} > 0.5\tau_{p,i_2}(3/\tau_{em} - 1/\tau_{eM}), \tag{9.62}$$

$$a_{p,i_2}^{\max}/a_{em}^{\max} > \tau_{eM}/\tau_{em}, \tag{9.63}$$

$$a_{p,i_2}^{\max}/a_{eM}^{\max} > 0.5\tau_{p,i_2}(1/\tau_{em} + 1/\tau_{eM}). \tag{9.64}$$

Then by virtue of Shinar et al. (2012), the AG2 optimal strategies are

$$\bar{u}^*(\vartheta, z_c) = \bar{v}^*(\vartheta, z_c) = \begin{cases} \text{arbitrary}, & (\vartheta, z_c) \in \bar{D}_0 \\ \\ \operatorname{sign} z_c, & (\vartheta, z_c) \notin \bar{D}_0, \end{cases} \tag{9.65}$$

where $\bar{D}_0$, called in Shinar et al. (2012) the singular zone, is given as

$$\bar{D}_0 = \Big\{(\vartheta, z_c) : \ \vartheta \in (0, \vartheta_0], \ \ z_c^-(\vartheta, \vartheta_0; s_e, \tilde{x}_{30}) < z_c < z_c^+(\vartheta, \vartheta_0; s_e, \tilde{x}_{30})\Big\}, \tag{9.66}$$

$$z_c^+(\vartheta, \vartheta_0; s_e, \tilde{x}_{30}) \triangleq \int_0^{\vartheta} R(\xi, \vartheta_0; \omega_p, \omega_e^1, \omega_e^2, s_e, \tilde{x}_{30}) d\xi > 0, \ \ \vartheta > 0, \tag{9.67}$$

$$z_c^-(\vartheta, \vartheta_0; s_e, \tilde{x}_{30}) \triangleq -\int_0^{\vartheta} R(\xi, \vartheta_0; \omega_p, \omega_e^1, \omega_e^2, s_e, -\tilde{x}_{30}) d\xi < 0, \ \ \vartheta > 0, \tag{9.68}$$

$$R(\vartheta, \vartheta_0; \omega_p, \omega_e^1, \omega_e^2, s_e, \nu) \triangleq R(\vartheta, \vartheta_0) = \begin{cases} R_1(\vartheta, \vartheta_0), \ \vartheta \in [\vartheta_e, \vartheta_0], \\ \\ R_2(\vartheta, \vartheta_0), \ \ \vartheta \in [0, \vartheta_e), \end{cases} \tag{9.69}$$

$$R_1(\vartheta, \vartheta_0) = R_1(\vartheta, \vartheta_0; \omega_p, \omega_e^{j_{e1}}) \triangleq h(\vartheta, \tau_p, a_p^{\max}) - h_e(\vartheta, \tau_{e,j_{e1}}, a_{e,j_{e1}}^{\max}) - h_x(\vartheta, \tau_{e,j_{e1}})\Big(-\nu \exp((\vartheta - \vartheta_0)/\tau_{e,j_{e1}}) + G_1(\vartheta)\Big), \tag{9.70}$$

$$R_2(\vartheta, \vartheta_0) = R_2(\vartheta, \vartheta_0; \omega_p, \omega_e^{j_{e1}}, \omega_e^{j_{e2}}) \triangleq h(\vartheta, \tau_p, a_p^{\max}) - h_e(\vartheta, \tau_{e,j_{e2}}, a_{e,j_{e2}}^{\max}) - h_x(\vartheta, \tau_{e,j_{e2}})\Big(-\nu \exp\Big[(\vartheta - \vartheta_e)/\tau_{e,j_{e2}} + (\vartheta_e - \vartheta_0)/\tau_{e,j_{e1}}\Big] + G_2(\vartheta) + G_3(\vartheta)\Big), \tag{9.71}$$

$$G_1(\vartheta) \triangleq a_{e,j_{e1}}^{\max}\Big(\exp((\vartheta - \vartheta_0)/\tau_{e,j_{e1}}) - 1\Big), \tag{9.72}$$

$$G_2(\vartheta) \triangleq G_1(\vartheta_e) \exp((\vartheta - \vartheta_e)/\tau_{e,j_{e2}}), \tag{9.73}$$

$$G_3(\vartheta) \triangleq a_{e,j_{e2}}^{\max}\Big(\exp((\vartheta - \vartheta_e)/\tau_{e,j_{e2}}) - 1\Big). \tag{9.74}$$

The value of AG2 is

$$\bar{J}^* = \bar{J}^*(\vartheta_p, w_{2p}) = \begin{cases} C_2 = \text{const}, & (\vartheta_p, w_{2p}) \in \bar{D}_0, \\ |w_{2p}| + \int\limits_{\vartheta_p}^{0} H_2(\xi)d\xi, & (\vartheta_p, w_{2p}) \notin \bar{D}_0, \end{cases} \tag{9.75}$$

where

$$H_2(\vartheta, s_p) \triangleq h(\vartheta, \tau_{p,i_2}, a_{p,i_2}^{\max}) - h_{e2}(\vartheta, \tau_e, a_e^{\max}). \tag{9.76}$$

Remark 7. For the sake of simplicity, we use in the further analysis the bang-bang strategies (9.65) in the entire AG2 space.

9.5.3 Optimal Pursuer Schedule

The AG2 value (9.75) depends on the dynamic schedules s_p and s_e of the pursuer and the evader: $\bar{J}^* = \bar{J}^*(s_p, s_e)$. In order to obtain the minimal value of $|z_c(0)|$ against the worst case evader control, the pursuer should choose the schedule as the solution of the following min-max optimization problem:

$$s_p^* = \arg \min_{s_p \in \mathcal{S}_p} \max_{s_e \in \mathcal{S}_e} \bar{J}^*(s_p, s_e). \tag{9.77}$$

The corresponding worst case (from the pursuer's viewpoint) evader's schedule is

$$s_e^* = \arg \max_{s_e \in \mathcal{S}_e} \bar{J}^*(s_p^*, s_e). \tag{9.78}$$

Based on the schedules s_p^* and s_e^*, the hybrid robust capture zone is constructed as

$$\mathcal{C}_p = \left\{(\bar{\vartheta}_0, \bar{z}_0) : \ \bar{\vartheta}_0 \in [0, \vartheta_0], |\bar{z}_0| \leq \mathcal{Z}_p(\bar{\vartheta}_0)\right\}, \tag{9.79}$$

where

$$\mathcal{Z}_p(\bar{\vartheta}_0) = \max\left\{\bar{z}_0 : \ \bar{J}^*(s_p^*, s_e^*) = 0\right\}. \tag{9.80}$$

9.5.4 Numerical Example

In this example, the hybrid robust capture zone $\mathcal{C}_p$ is constructed for the dynamics data, presented in Table 9.1, and $\vartheta_0 = 4$ s.

Table 9.1 Players' dynamics data

	Time constant (s)	Max. acc.command (m/s^2)
Pursuer	$\tau_{p1} = 0.1$	$a_{p1}^{\max} = 150$
	$\tau_{p2} = 0.8$	$a_{p2}^{\max} = 230$
Evader	$\tau_{e1} = 0.4$	$a_{e1}^{\max} = 100$
	$\tau_{e2} = 0.5$	$a_{e2}^{\max} = 120$

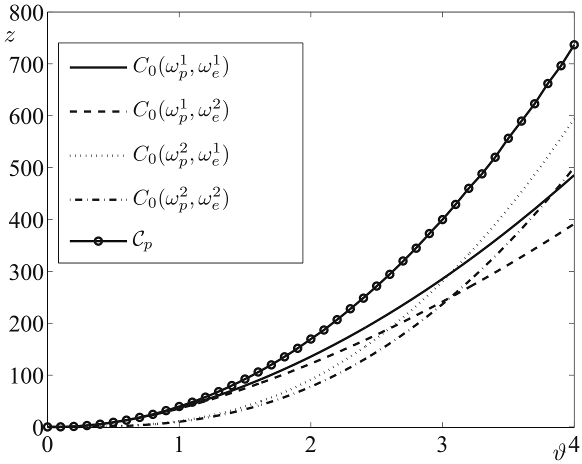

Fig. 9.3 Robust capture zones comparison

In the two following figures the characteristics of the hybrid robust capture zone are illustrated. In Fig. 9.3, it is shown that the hybrid robust capture zone is larger then the union of all the fixed dynamics robust capture zones:

$$\mathcal{C}_p \supset \bigcup_{i,j=1}^{2} C_0(\omega_p^i, \omega_e^j). \tag{9.81}$$

In Fig. 9.4, three optimal ZEM trajectories, emanating from the upper bound of the hybrid robust capture zone for $\bar{\vartheta}_0 = 4$ s, $\bar{\vartheta}_0 = 3.1$ s and $\bar{\vartheta}_0 = 2.1$ s, are depicted. These trajectories are generated by the optimal schedules, defined in (9.77) and (9.78), respectively, by using the AG1 and AG2 optimal strategies

$$u^* = v^* = \begin{cases} \operatorname{sign}(w(\vartheta)) & \vartheta_0 \geq \vartheta > \vartheta_p, \\ \\ \operatorname{sign}(z_c(\vartheta)) & \vartheta_p \geq \vartheta > 0. \end{cases} \tag{9.82}$$

They show the discontinuities created by the "jumps" and demonstrate that along all of them capture is achieved.

The schedules s_p^* and s_e^* for different values of $\bar{\vartheta}_0$ are presented in Table 9.2.

From Table 9.2, we can observe that for all $\bar{\vartheta}_0$, the optimal order (i_1^*, i_2^*) of the pursuer's dynamic modes is the same as in the differential game with a hybrid

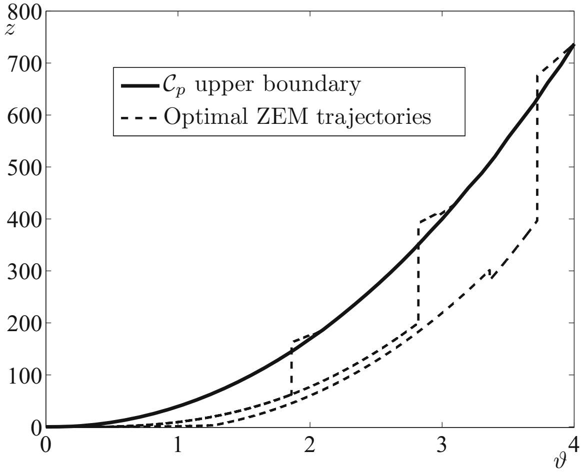

Fig. 9.4 Optimal ZEM trajectories

Table 9.2 Schedules s_p^* and s_e^*

$\bar{\vartheta}_0$ (s)	$\bar{z}_0$ (m)	s_p^*			s_e^*		
		i_1^*	i_2^*	ϑ_p^* (s)	j_1^*	j_2^*	ϑ_e^* (s)
4.0	736.82	1	2	3.72	1	2	3.36
3.9	696.41	1	2	3.58	1	2	3.14
3.8	662.09	1	2	3.56	2	1	0
3.7	623.11	1	2	3.46	1	2	3.50
3.6	589.64	1	2	3.36	2	1	0
3.5	556.51	1	2	3.22	1	2	3.46
3.4	519.90	1	2	3.12	1	2	3.36
3.3	487.99	1	2	3.02	2	1	0.02
3.2	459.97	1	2	2.96	1	2	2.80
3.1	428.96	1	2	2.82	1	2	2.98
3.0	399.71	1	2	2.76	1	2	2.76
2.9	372.88	1	2	2.66	1	2	2.78
2.8	344.57	1	2	2.56	1	2	2.68
2.7	319.96	1	2	2.42	1	2	2.62
2.6	294.16	1	2	2.36	2	1	0.04
2.5	271.69	1	2	2.22	1	2	2.38
2.4	248.29	1	2	2.16	1	2	2.16
2.3	227.89	1	2	2.06	2	1	0.06
2.2	206.80	1	2	1.92	1	2	2.04
2.1	186.90	1	2	1.86	1	2	2.02

Table 9.2 Schedules s_p^* and s_e^* (contd.)

$\bar{\vartheta}_0$ (s)	$\bar{z}_0$ (m)	s_p^*			s_e^*		
		i_1^*	i_2^*	ϑ_p^* (s)	j_1^*	j_2^*	ϑ_e^* (s)
2.0	169.51	1	2	1.76	1	2	1.76
1.9	151.77	1	2	1.62	1	2	1.62
1.8	135.13	1	2	1.52	1	2	1.68
1.7	119.57	1	2	1.46	1	2	1.46
1.6	105.95	1	2	1.36	1	2	1.52
1.5	92.37	1	2	1.26	1	2	1.46
1.4	79.78	1	2	1.12	1	2	1.24
1.3	68.17	1	2	1.02	1	2	1.06
1.2	57.53	1	2	0.96	1	2	1.04
1.1	47.81	1	2	0.86	1	2	0.94
1.0	39.39	1	2	0.76	1	2	0.92
0.9	31.43	1	2	0.62	1	2	0.74
0.8	24.37	1	2	0.56	2	1	0.08
0.7	18.39	1	2	0.42	2	1	0.02
0.6	13.20	1	2	0.36	1	2	0.52
0.5	8.91	1	2	0.26	1	2	0.42
0.4	5.41	1	2	0.12	1	2	0.32
0.3	2.76	1	2	0.02	1	2	0.22
0.2	1.02	1	2	0	1	2	0.12
0.1	0.18	1	2	0.02	1	2	0.02

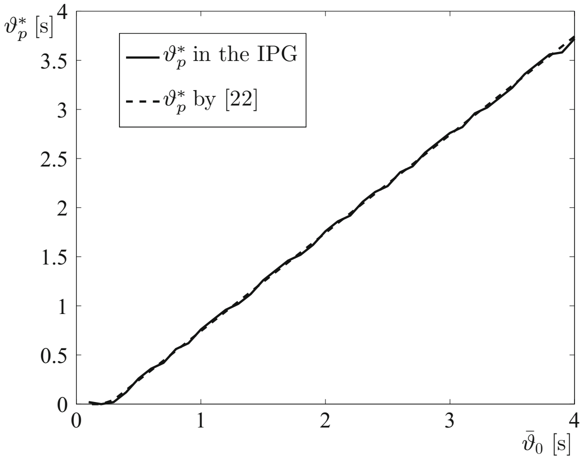

Fig. 9.5 Optimal switch moments ϑ_p^*

dynamics pursuer and a fixed dynamics evader (Shinar et al. 2009b). Namely, the pursuer switches from a smaller value of τ_p to a larger ($\tau_{p1} = 0.1$ s, $\tau_{p2} = 0.8$ s in the example). Moreover, the optimal pursuer's switch moment ϑ_p^* is very close to the value calculated in Shinar et al. (2009b) as

$$\tilde{\vartheta}_p^* = \bar{\vartheta}_0 - \tau_{p1} \ln \left[\frac{a_{p1}^{\max}(\tau_{p2} - \tau_{p1})}{\tau_{p1}(a_{p2}^{\max} - a_{p1}^{\max})} \right]. \tag{9.83}$$

The values of ϑ_p^*, obtained in the numerical example as functions of $\bar{\vartheta}_0$, are compared in Fig. 9.5 to the analytical results of Shinar et al. (2009b), indicating almost coincidence, taking into account the limited numerical accuracy.

In the set of the optimal (worst case) evader's schedules s_e^*, two subsets can be distinguished. The first subset corresponds to the seven values $\bar{\vartheta}_0 = 0.7, 0.8, 2.3, 2.6, 3.3, 3.6$ and 3.8 s. In this subset, the evader's switch moment ϑ_e^* is zero or very close to zero, while the order (j_1^*, j_2^*) is (2, 1), meaning that the evader actually do not change its dynamic mode and employs the second mode during the entire engagement. The rest of the schedules s_e^* are close to the optimal evader's schedules in the differential game with a fixed dynamics pursuer and a hybrid dynamics evader (Shinar et al. 2009a): switch from a smaller τ_e to a larger ($\tau_{e1} = 0.4$ s, $\tau_{e2} = 0.5$ s in the example) and calculating $\vartheta_e^* = \tilde{\vartheta}_e^*$ by replacing the subscript "*p*" to "*e*" in (9.83). The values of ϑ_e^* from the second subset and the respective values $\tilde{\vartheta}_e^*$ as functions of $\bar{\vartheta}_0$ are compared in Fig. 9.6, indicating a reasonable good matching, taking into account the limited numerical accuracy.

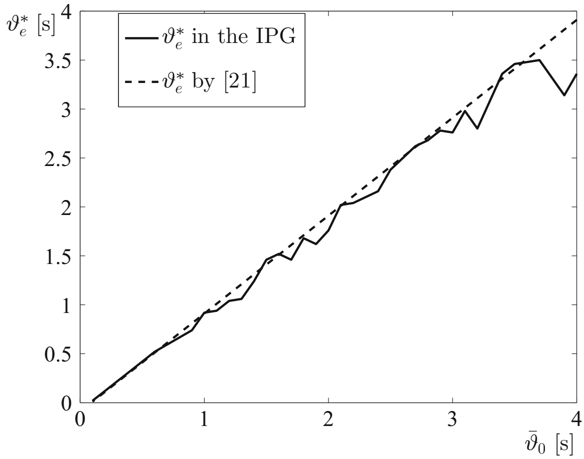

Fig. 9.6 Optimal switch moments ϑ_e^*

9.6 Conclusions

In this paper, a pursuit-evasion differential game of kind where both players have hybrid dynamics was considered. Each player has two possible fixed dynamic modes at its disposal and can switch from one mode to the other once during the game. The cost functional of the associated game of degree is the miss distance, to be minimized by the pursuer and maximized by the evader. This game was solved from the pursuer's viewpoint, thus called the Hybrid Pursuit Game (HPG). In the HPG, the pursuer knows perfectly the current state vector and the set of the possible evader's dynamic modes, but not the current mode. Thus, the HPG is an incomplete information game. By introducing a new aiming point of the pursuer (the center of the convex hull of the uncertainty set), the HPG was converted to a new complete information reduced dimension differential game, the Reduced Hybrid Pursuit Game (RHPG). In the RHPG, a state jump occurs at the moment when the pursuer changes its dynamic mode, introducing an impulsive character to the dynamics.

The RHPG was solved in several stages.

1. Assuming a fixed order of the pursuer's dynamics and a fixed switch moment of the pursuer's dynamics modes (constituting a fixed dynamic schedule of the pursuer), an auxiliary differential game (AG1) was formulated on the time interval between the engagement beginning and the switch moment of the pursuer's dynamics. This game, where the pursuer uses its first dynamic mode and the cost functional is the zero effort miss distance immediately after the switch, was solved by its scalarization.

2. A second auxiliary differential game (AG2) was formulated and solved on the time interval between the switch moment of the pursuer's dynamics and the end of the engagement. In this game, the pursuer uses its second dynamic mode.
3. The outcome of the second auxiliary differential game is maximized, for a given fixed pursuer's dynamic schedule, over all possible evader's dynamic schedules (i.e. over the two possible orders of the evader's dynamic modes and its all possible dynamics switch moments). The optimal (worst case) evader's extended strategy is composed of the evader's strategy and the maximizing evader's dynamic schedule against the given pursuer's strategy. This stage yields the guaranteed RHPG value for a given pursuer's schedule.
4. The next step is to find the pursuer's dynamic schedule, minimizing the guaranteed RHPG value. This minimal guaranteed value is the upper value of the RHPG. The minimizing pursuer's schedule, along with the pursuer's strategy constructed at the first two stages, forms the pursuer's extended optimal strategy.
5. Based on the solution of the RHPG, the (robust) capture zone of the game of kind is constructed, as the set of all initial positions, for which the RHPG upper value is zero.

In the numerical example, included in the paper, it is shown that the robust capture zone of the HPG is larger than the union of all possible robust capture zones, constructed for the four possible robust capture zones with fixed dynamics. The example also shows that the optimal dynamic schedule of the pursuer coincides with the optimal pursuer schedule in the differential game between a pursuer with hybrid dynamics and an evader with fixed dynamics, solved previously by the authors. Also, the optimal (worst case) dynamic schedule of the evader in the majority of the samples of the example is rather close to the optimal dynamic schedule of the evader in the differential game between a pursuer with fixed dynamics and an evader with hybrid dynamics. In the few other cases the evader's dynamics is not changed during the game. These observations seem to be logical, but will need a rigorous justification in a future work.

References

Bartolini G, Zolezzi T (1986) Control of nonlinear variable structure systems. J Math Anal Appl 118:42–62

Chen X, Fukuda T (1997) Variable structure system theory based disturbance identification and its applications. Int J Control 68:373–384

Chernousko FL, Melikyan AA (1975) Some differential games with incomplete information. In: Lecture notes in computer science, vol 27. Springer, Berlin/Heidelberg, pp 445–450

Choi H (2004) An analysis and design method for uncertain variable structure systems with bounded controllers. IEEE Trans Autom Control 26:1139–1144

Gao Y, Lygeros J, Quincampoix M (2007) On the reachability problem for uncertain hybrid systems. IEEE Trans Autom Control 52:1572–1586

Glizer VY, Turetsky V (2008) Complete solution of a differential game with linear dynamics and bounded controls. Appl Math Res Express. Article ID abm012, 49 pp.

Grigorenko NL (1991) Game-theoretic problems of control with variable structure. Mosc Univ Comput Math Cybern 4:4–14

Gutman S (1979) On optimal guidance for homing missiles. J Guid Control Dyn 2:296–300

Gutman S (2006) Applied min-max approach to missile guidance and control. In: Progress in astronautics and aeronautics, vol 209. AIAA, Reston

Gutman S, Leitmann G (1976) Optimal strategies in the neighborhood of a collision course. AIAA J 14:1210–1212

Ho YC, Bryson AE, Baron S (1965) Differential games and optimal pursuit-evasion strategies. IEEE Trans Autom Control 10:385–389

Krasovskii N (1984) Problem on control in conditions of incomplete information. J Appl Math Mech 48:375–379

Kumkov S, Patsko V (1995) Optimal strategies in a pursuit problem with incomplete information. J Appl Math Mech 59:75–85

Lee Y, Kouvaritakis B (2000) Robust receding horizon predictive control for systems with uncertain dynamics and input saturation. Automatica 36:1497–1504

Mitchell IM, Bayen AM, Tomlin CJ (2001) Validating a Hamilton-Jacobi approximation to hybrid system reachable sets. In: Benedetto MDD, Sangiovanni-Vincentelli A (eds) Hybrid systems: computation and control. Lecture notes in computer science, vol 2034. Springer, New York, pp 418–432

Mitchell IM, Bayen AM, Tomlin CJ (2005) A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. IEEE Trans Autom Control 50:947–957

Petrosjan L (1993) Differential games of pursuit. World Scientific, Singapore

Riedinger P, Iung C, Kratz F (2003) An optimal control approach for hybrid systems. Eur J Control 9:449–458

Shima T, Shinar J (2002) Time-varying linear pursuit-evasion game models with bounded controls. J Guid Control Dyn 25:425–432

Shinar J (1981) Solution techniques for realistic pursuit-evasion games. In: Leondes C (ed) Advances in control and dynamic systems, vol 17. Academic, New York, pp 63–124

Shinar J, Glizer VY, Turetsky V (2009a) A pursuit-evasion game with hybrid evader dynamics. In: Proceedings of the European control conference, Budapest, pp 121–126

Shinar J, Glizer VY, Turetsky V (2009b) A pursuit-evasion game with hybrid pursuer dynamics. Eur J Control 15:665–684

Shinar J, Glizer VY, Turetsky V (2009c) Pursuit of an evader with hybrid dynamics. In: Preprints of the 3rd IFAC conference on analysis and design of hybrid systems, Zaragoza, pp 98–103

Shinar J, Glizer VY, Turetsky V (2010) Robust pursuit of a hybrid evader. Appl Math Comput 217:1231–1245

Shinar J, Glizer VY, Turetsky V (2012) Complete solution of a pursuit-evasion differential game with hybrid evader dynamics. Int Game Theory Rev 14:1250-014-1–1250-014-31

Shinar J, Glizer VY, Turetsky V (2013) The effect of pursuer dynamics on the value of linear pursuit-evasion games with bounded controls. In: Krivan V, Zaccour G (eds) Advances in dynamic games: theory, applications and numerical methods. Annals of the International Society of Dynamic Games, vol 13. Birkhauser, Cham, pp 313–350

Sussmann HJ (1999) A nonsmooth hybrid maximum principle: stability and stabilization of nonlinear systems. In: Lecture notes in control and information science, Springer, Berlin, vol 246, pp 325–354

Turetsky V, Shinar J (2003) Missile guidance laws based on pursuit-evasion game formulations. Automatica 39:607–618

Utkin VI (1983) Variable structure systems: present and future. Autom Remote Control 44:1105–1120

Chapter 10
A Double-Sided Jamming Game with Resource Constraints

Sourabh Bhattacharya, Ali Khanafer, and Tamer Başar

Abstract In this article, we study the problem of power allocation in teams of mobile agents in a conflict situation. Each team consists of two agents who try to split their available power between the tasks of communication and jamming the nodes of the other team. The agents have constraints on their total energy and instantaneous power usage. The cost function is the difference between the rates of erroneously transmitted bits of each team. We present a 2-level game formulation: At the higher level, the agents solve a continuous-kernel power allocation game at each instant. Based on the communications model, we present sufficient conditions on the physical parameters of the agents for the existence of a Pure Strategy Nash Equilibrium for the continuous-kernel power allocation game. At the lower level, we have a zero-sum differential game between the two teams and use *Isaacs'* approach to obtain necessary conditions for the optimal trajectories. The optimal power allocation scheme obtained at the upper level is used to solve the lower level differential game. This gives rise to a *games-in-games* scenario which is one of the first such phenomena documented in the literature.

Keywords Pursuit-evasion games • Jamming • Games-in-games • Nash equilibrium • Multi-level games

Math Classification Number(2010): 49N90

S. Bhattacharya (✉)
Department of Mechanical Engineering, Iowa State University, 2025 Black Engineering Building, Ames, IA 50011, USA
e-mail: sbhattac@iastate.edu

A. Khanafer • T. Başar
Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 West Main, Urbana, IL 61801, USA
e-mail: khanafe2@illinois.edu; basar1@illinois.edu

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_10

10.1 Introduction

The decentralized nature of wireless ad hoc networks makes them vulnerable to security threats. A prominent example of such threats is jamming: a malicious attack whose objective is to disrupt the communication of the victim network intentionally, causing interference or collision of packets at the receiver side. Jamming attack is a well-studied and active area of research in wireless networks. Unauthorized intrusion of such kind has started a race between the network operators and the hackers; accordingly, we have been witnessing a surge of new smart systems aiming to secure modern instrumentation and software from unwanted exogenous attacks.

Many defense strategies have been proposed by researchers against jamming in wireless networks. A brief survey of various techniques in jamming relevant to our research is provided in Bhattacharya and Başar (2010b). In the past, networks with multiple attackers have also been considered in the literature. In Han et al. (2009a,b), the authors consider the interaction between a source-destination pair, an eavesdropper, and friendly jammers. The source can buy "jamming power" from the friendly jammers so as to disguise the eavesdropper. This allows the source to achieve increased secrecy rate. In these papers, the authors study the problem in the context of a Stackelberg game and show that a trade-off exists between the price announced by the jammers and the resulting performance. A similar problem was tackled in Dong et al. (2010) where relay nodes can help the source in the presence of multiple eavesdroppers. The authors of Dong et al. (2010) propose different relaying schemes and study two design problems: minimizing the transmission power subject to a minimum secrecy rate and maximizing the secrecy rate subject to a total power constraint. Analysis shows that relaying yields improved performance when compared to direct transmission in malicious environments. Different from the aforementioned references, our work here considers non-friendly jamming teams, i.e., the security bottleneck considered here is jamming and not eavesdropping. Moreover, it appears that pursuit-evasion strategies for jamming teams have not been studied before.

In the past, we have analyzed various scenarios of evading jamming attacks among autonomous agents. In the case of a *single* jammer trying to intrude the communication link between a transmitter and a receiver, the problem can be formulated as a multiplayer (specifically, three-player) pursuit-evasion game (Isaacs 1965; Başar and Olsder 1999). In Bhattacharya and Başar (2010b), we investigated the problem of finding motion strategies for two unmanned autonomous vehicles (UAVs) to evade jamming in the presence of an aerial intruder. We considered a differential game theoretic approach to compute optimal strategies by a team of UAVs. We formulated the problem as a zero-sum pursuit-evasion game. The cost function was picked as the termination time of the game. We used *Isaacs'* approach to derive necessary conditions to arrive at the equations governing the saddle-point strategies of the players. In Bhattacharya and Başar (2010d), we extended the previous analysis to a team of heterogeneous vehicles containing UAVs and autonomous ground vehicles (AGVs). In Bhattacharya and Başar (2010a), we analyzed the problem of multiple jammers intruding the communication network in a formation of UAVs. In Bhattacharya and Başar (2010c), we analyzed the problem of connectivity maintenance in multi-agent systems in the presence of a jammer. In this current work, we study a scenario where a *team* of malicious nodes launches a jamming attack on another team, which is capable of jamming as well. Our analysis takes into consideration constraints on energy and power among the agents. Moreover, we relate the problem of optimal power allocation for communication and jamming to the communication model between the agents. Finally, we provide a sufficient condition for the existence of an optimal decision strategy among the agents based on the physical parameters of the problems.

The rest of the article is organized as follows. We formulate the problem in Sect. 10.2 and explain the underlying notation. In Sect. 10.3, we introduce and solve an associated optimal control problem. The Nash equilibrium properties of the team power control problem are studied in Sect. 10.4, and the specific example of systems employing uncoded M-quadrature amplitude modulations (QAM) follows in Sect. 10.5. We conclude the paper and provide future directions in Sect. 10.6.

10.2 Problem Formulation

Consider two teams of mobile agents. Each agent (synonymously, player) is communicating with members of the team it belongs to and, at the same time, jamming the communication between members of the other team. We consider a scenario where each team has two members, though at a conceptual level our development applies to teams comprised of more than two players as well. Team A is comprised of the two players $\{1^a, 2^a\}$ and Team B is comprised of the two players $\{1^b, 2^b\}$. The players move on a plane and therefore, have two degrees of freedom (x, y). The dynamics of the players are given by the following equations:

- Team A:

$$\left.\begin{aligned} \dot{x}_i^a &= f_{x_i}^a(\mathbf{x_i^a}, \mathbf{u_i^a}, t) \\ \dot{y}_i^a &= f_{y_i}^a(\mathbf{x_i^a}, \mathbf{u_i^a}, t) \end{aligned}\right\} \; i \in \{1, 2\} \tag{10.1}$$

- Team B:

$$\left.\begin{aligned} \dot{x}_i^b &= f_{x_i}^b(\mathbf{x_i^b}, \mathbf{u_i^b}, t) \\ \dot{y}_i^b &= f_{y_i}^b(\mathbf{x_i^b}, \mathbf{u_i^b}, t) \end{aligned}\right\} \; i \in \{1, 2\} \tag{10.2}$$

In the above equations, $\mathbf{x}_i$ and $\mathbf{u}_i$ denote vectors representing the state and control input of agent $\mathbf{i}$, with the superscript (a or b) identifying the corresponding team. The state space of the entire system is represented by $\mathbf{X} \simeq \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$. Moreover, $\mathbf{u}_i \in \mathscr{U}_i \simeq \{\phi : [0, t] \to \mathscr{A}_i \mid \phi(\cdot) \quad \text{is measurable}\}$, where $\mathscr{A}_i \subset \mathbb{R}^{p_i}$ and $f : \mathbb{R}^2 \times \mathscr{A}_i \times \mathbb{R} \to \mathbb{R}$ is uniformly continuous, bounded and Lipschitz continuous in $\mathbf{x_i}$ for each fixed $\mathbf{u_i}$. Consequently, given a fixed $\mathbf{u_i}(\cdot)$ and an initial point, there exists a unique trajectory solving (10.1) and (10.2) (Arnold 1983).

Next, we describe the physical layer communications model in the presence of a jammer, which is motivated by Tague et al. (2009). For each transmitter and receiver pair, we assume a free space path loss (FSPL) model. In telecommunication, FSPL is the loss in signal strength of an electromagnetic wave that would result from a line-of-sight path through free space (usually air), with no obstacles nearby to cause reflection or diffraction. Given that the transmitter and the receiver are separated by a distance d, and the transmitter transmits with constant power P_T, the received signal power P_R is given by

$$P_R = \rho P_T (1 + d)^{-\alpha}, \tag{10.3}$$

where ρ depends on the antennas' gains and, under the FSPL model, is given by:

$$\rho = \frac{G_T G_R \lambda^2}{(4\pi)^2},$$

where λ is the signal's wavelength; G_T and G_R are the transmit and receive antennas' gains, respectively, in the line of sight direction; and α is the path loss exponent, whose value is normally in the range of 2–4 (where 2 corresponds to propagation in free space, and corresponds to 4 relatively lossy environments and for the case of full specular reflection from the earth surface, the so-called Flat Earth model). In real scenarios, ρ is very small in magnitude. For example, using nondirectional antennas and transmitting at 900 MHz, we have $\rho = \frac{1 \cdot 1 \cdot 0.33}{(4\pi)^2} = 6.896 \times 10^{-4}$.

The signal-to-interference plus noise ratio (SINR) s is given by

$$s = \frac{P_R}{I + \sigma}, \tag{10.4}$$

where I is the interference level and σ is the ambient noise level. The Bit Error Rate (BER) is given by the following expression:

$$p(t) = g(s), \tag{10.5}$$

where $g(\cdot)$ is a decreasing function of s. Explicit expressions for $g(\cdot)$ are provided in Sect. 10.5 where we consider the example of M-QAM. Each player uses its power for the following purposes: (1) communicating with the team-mate, and (2) jamming the communication of the other team. We assume that the frequencies at which agents within Team A and Team B communicate among themselves are different.

Let $P_i^a(t)$ and $P_i^b(t)$ denote the instantaneous power levels for communication used by player i in Team A and Team B, respectively. Since the agents are mobile, there are limitations on the amount of energy available to each agent, which are dictated by the capacity of the power source carried by each agent. We model this restriction as the following integral constraint for each agent:

$$\int_0^T P_i(t)dt \leq E. \tag{10.6}$$

The game is said to terminate when any one agent runs out of power, that is (6) is violated for the first time.

In addition to the energy constraints, there are limitations on the maximum power level of the devices that are used onboard by each agent for the purpose of communication. For each player, this constraint is modeled by the following set of inequalities:

$$0 \leq P_i^a(t), P_i^b(t) \leq P_{max}. \tag{10.7}$$

For an initial position $\mathbf{x}_0 \in \mathbf{X}$, the outcome of the game π, is given by the following expression:

$$\pi(\mathbf{x}_0, \mathbf{u}_1^a, \mathbf{u}_2^a, \mathbf{u}_1^b, \mathbf{u}_2^b) = \int_0^T \underbrace{[p_1^a(t) + p_2^a(t) - p_1^b(t) - p_2^b(t)]}_{L} dt,$$

where $p_i^a(t)$ and $p_i^b(t)$ are the bit error rates (BERs) for agent i in team A and team B, respectively, and T is the time of termination of the game. The function p_i depends on s_i, i.e, the SINR perceived by agent i. From (10.3), s_i depends on the mutual distances between the players. Therefore, we can see that the outcome functional, π, depends on the states of the players and hence, their control inputs. The outcome functional models the difference in the erroneous communication packets exchanged between the members of the same team during the entire course of the game. The objective of team A is to minimize π and the objective of team B is to maximize it.

At every instant, each agent has to decide on the fraction of the power that needs to be allocated for communication and jamming. Let γ denote the variable that represents the power allocation by a member of Team A. γ_j^i denotes the fraction of power allocated by agent i in Team A for jamming the communication signal received by agent j in Team B. γ^{ij} denotes the fraction of power allocated by agent i in Team A to communicate with agent j in Team A. Similarly, δ denotes the variable that represents the power allocation by a member of Team B. δ_i^j denotes the fraction of power allocated by agent i in Team B for jamming the communication signal received by agent j in Team A. δ_{ij} denotes the fraction of power allocated by agent i in Team B to communicate with agent j in Team B. Each decision variable is a non-negative real number and lies in the interval [0, 1].

The variable d is used to represent the distance between two agents. d_j^i represent the distances between agent i in Team A and agent j in Team B. d^{ij} and d_{ij} represent the distances between two agents in Team A and Team B, respectively.

Table 10.1 provides a list of decision variables for the players that models this allocation. The decision variables belonging to each row add up to one. The fraction of the total power allocated by the player in row i to the player in column j is given

Table 10.1 Decision variables and distances among agents

	1^b	2^b	1^a	2^a
1^a	γ_1^1, d_1^1	γ_2^1, d_2^1		γ^{12}, d^{12}
2^a	γ_1^2, d_1^2	γ_2^2, d_2^2	γ^{21}, d^{21}	
1^b		δ_{12}, d_{12}	δ_1^1, d_1^1	δ_1^2, d_1^2
2^b	δ_{21}, d_{21}		δ_2^1, d_2^1	δ_2^2, d_2^2

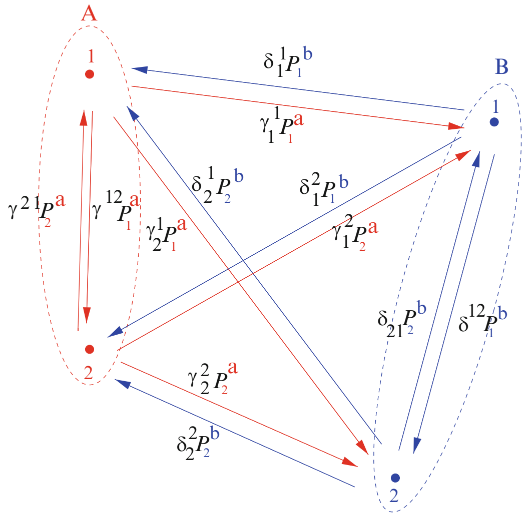

Fig. 10.1 Power allocations among the agents for communication as well as jamming

by the first entry in the cell (i, j). This allocated power is used for jamming if the player in column j belongs to the other team; otherwise, it is used to communicate with the agent in the same team. Similarly, the distance between the agent in row i and the agent in column j is given by the second entry in cell (i, j). Since distance is a symmetric quantity, $d^{ij} = d^{ji}$ and $d_{ij} = d_{ji}$. Figure 10.1 summarizes the power allocations between the members of the same team as well as between the members of different teams.

In the above game, each agent has to compute the following variables at every instant:

1. The instantaneous control, $\mathbf{u_i}(t)$.
2. The instantaneous power level, $P_i(t)$.
3. All the decision variables present in the row corresponding to the agent in Table 10.1.

In the next section, we analyze the problem of computing the optimal controls for each agent.

10.3 Optimal Control Problem

From the problem formulation presented in the previous section, we can conclude that the objective functions of the two teams are in conflict. The tuple $(\mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*})$ is said to be optimal (or, in pair-wise saddle-point equilibrium) for the players if it satisfies the following pair of inequalities:

$$\pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b}, \mathbf{u}_2^{b}] \leq \pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \tag{10.8}$$

$$\pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \leq \pi[\mathbf{x}_0, \mathbf{u}_1^{a}, \mathbf{u}_2^{a}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \tag{10.9}$$

In simple terms, the above equations imply that agents in Team A are solving a joint optimization problem of minimizing the outcome. Similarly, agents in Team B are solving a joint optimization problem of maximizing the outcome. Moreover, the two teams are playing a zero-sum game against one another. In this case, the value of the game, denoted by the function $J : \mathbf{X} \to \mathbb{R}$, can be defined as follows:

$$J(\mathbf{x}_0) = \pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \tag{10.10}$$

The value of the game is unique at a point $\mathbf{x}_0$ in the state-space. An important property satisfied by the value of the game is the *Nash equilibrium* property. The tuple $(\mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*})$ is said to be in Nash equilibrium if no unilateral deviation in strategy by a player can lead to a better outcome for that player. Hence, there is no motivation for the players to deviate from their equilibrium strategies. In terms of the outcome of the game, the strategies $(\mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*})$ are in Nash equilibrium (for the 4-player game) if they satisfy the following 4-tuple of inequalities:

$$\left.\begin{array}{l}\pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b}] \\ \pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b}, \mathbf{u}_2^{b*}]\end{array}\right\} \leq \pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}]$$

$$\pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \leq \left\{\begin{array}{l}\pi[\mathbf{x}_0, \mathbf{u}_1^{a}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \\ \pi[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}]\end{array}\right. \tag{10.11}$$

In general, there may be multiple sets of strategies for the players that are in Nash equilibrium. Assuming the existence of a value, as captured by (10.10), and the existence of a unique Nash equilibrium, we can conclude that the Nash equilibrium concept of person-by-person optimality given in (10.11) is a sufficient condition to be satisfied for the value of the game. In view of this, obtaining the set of strategies that are in Nash equilibrium yields the optimal strategies for the players. In the following analysis, we assume the aforementioned conditions in order to compute the optimal strategies.

The Hamiltonian of the system is given by the following expression:

$$\begin{aligned} H &= L + \nabla J \cdot f(\mathbf{x}) \\ &= p_1^a(t) + p_2^a(t) - p_1^b(t) - p_2^b(t) + \nabla J \cdot f(\mathbf{x}), \end{aligned} \tag{10.12}$$

where the superscript associated with p represents the team, and the subscript denotes the player. In order to compute the optimal control of the players, we will use the *Isaacs'* conditions (Isaacs 1965) which are the following:

1.

$$\left.\begin{array}{l} H[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b}] \\ H[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b}, \mathbf{u}_2^{b*}] \end{array}\right\} \leq H[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}]$$

$$H[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \leq \left\{\begin{array}{l} H[\mathbf{x}_0, \mathbf{u}_1^{a}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \\ H[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] \end{array}\right.$$

2. $H[\mathbf{x}_0, \mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}] = 0$

The agents in Team A want to minimize the Hamiltonian at every instant, and the agents in Team B want to maximize it. The dynamics of the agents are decoupled. Therefore, the Hamiltonian is separable in its controls and, hence, the order of taking the extrema becomes inconsequential. As a result, the optimal controls of the players are given by the following expression from Isaacs' first condition

$$(\mathbf{u}_1^{a*}, \mathbf{u}_2^{a*}, \mathbf{u}_1^{b*}, \mathbf{u}_2^{b*}) = \arg \max_{(\mathbf{u}_1^{b}, \mathbf{u}_2^{b})} \min_{(\mathbf{u}_1^{a}, \mathbf{u}_2^{a})} H$$

The optimal control $\mathbf{u}_i$ is obtained by the following expression:

$$\left.\begin{array}{l} \mathbf{u}_i^{a*} = \arg\min_{\mathbf{u}_i^a} \frac{\partial J}{\partial \mathbf{x}_i^a} \cdot f_i^a(\mathbf{x}_i^a, \mathbf{u}_i^a, t) \\ \mathbf{u}_i^{b*} = \arg\max_{\mathbf{u}_i^b} \frac{\partial J}{\partial \mathbf{x}_i^b} \cdot f_i^b(\mathbf{x}_i^b, \mathbf{u}_i^b, t) \end{array}\right\} \quad i = 1, 2$$

Additionally, the gradient of the value function satisfies the *retrogressive path equations* (RPEs) given by the following partial differential equation:

$$\frac{\partial \nabla J}{\partial \tau} = \frac{\partial H}{\partial \mathbf{x}},$$

where τ is the retrograde time or time left for termination.

Since termination is only a function of the power of each player, J is independent of the position of the players on the terminal manifold. Therefore, $\nabla J = 0$ at termination. This forms the boundary condition for the RPEs.

In the next section, we address the problem of power allocation.

10.4 Power Allocation

From (10.4), the SINR received by each agent in terms of the power levels of the other agents as well as their mutual distances is given by the following expressions

$$
\begin{aligned}
s_1^a &= \frac{P_2^a(t)\gamma^{21}(1+d^{12})^{-\alpha}}{\frac{\sigma}{\rho} + P_1^b(t)\delta_1^1(1+d_1^1)^{-\alpha} + P_2^b(t)\delta_2^1(1+d_2^1)^{-\alpha}} \\
s_2^a &= \frac{P_1^a(t)\gamma^{12}(1+d^{12})^{-\alpha}}{\frac{\sigma}{\rho} + P_1^b(t)\delta_1^2(1+d_1^2)^{-\alpha} + P_2^b(t)\delta_2^2(1+d_2^2)^{-\alpha}} \\
s_1^b &= \frac{P_2^b(t)\delta_{21}(1+d_{12})^{-\alpha}}{\frac{\sigma}{\rho} + P_1^a(t)\gamma_1^1(1+d_1^1)^{-\alpha} + P_2^a(t)\gamma_1^2(1+d_1^2)^{-\alpha}} \\
s_2^b &= \frac{P_1^b(t)\delta_{12}(1+d_{12})^{-\alpha}}{\frac{\sigma}{\rho} + P_1^a(t)\gamma_2^1(1+d_2^1)^{-\alpha} + P_2^a(t)\gamma_2^2(1+d_2^2)^{-\alpha}}
\end{aligned}
\tag{10.13}
$$

From the expression in (10.12), we can conclude that the power allocation among the agents only affects the term L in the Hamiltonian. Therefore, Isaacs' first condition leads to the following power allocation problem among the agents.

10.4.1 Team A

The objective of each agent is to minimize L.

1. Player 1:

$$
\min_{P_1^a,\gamma_1^1,\gamma_2^1,\gamma^{12}} L \Rightarrow \min_{P_1^a,\gamma_1^1,\gamma_2^1,\gamma^{12}} \underbrace{(p_2^a - p_1^b - p_2^b)}_{L_1^a}
\tag{10.14}
$$

   subject to:

$$
P_1^a(t) \le P_{\max}
$$

$$
\gamma_1^1 + \gamma_2^1 + \gamma^{12} = 1, \quad \gamma_1^1, \gamma_2^1, \gamma^{12} \ge 0 \Rightarrow \gamma \in \Delta^3,
$$

   where Δ^3 denotes the simplex in $\mathbb{R}^3$.

2. Player 2:

$$
\min_{P_2^a,\gamma_1^2,\gamma_2^2,\gamma^{21}} L \Rightarrow \min_{P_2^a,\gamma_1^2,\gamma_2^2,\gamma^{21}} \underbrace{(p_1^a - p_1^b - p_2^b)}_{L_2^a}
\tag{10.15}
$$

   subject to:

$$
P_2^a(t) \le P_{\max}
$$

$$
\gamma_1^2 + \gamma_2^2 + \gamma^{21} = 1, \quad \gamma_1^1, \gamma_2^1, \gamma^{21} \ge 0 \Rightarrow \gamma \in \Delta^3
$$

10.4.2 Team B

The objective of each agent is to maximize L.

1. Player 1:

$$\max_{P_1^b,\delta_1^1,\delta_2^1,\delta_{12}} L \Rightarrow \max_{P_1^b,\delta_1^1,\delta_2^1,\delta_{12}} \underbrace{(p_1^a + p_2^a - p_2^b)}_{L_1^b} \tag{10.16}$$

subject to:

$$P_b^1(t) \leq P_{\max}$$

$$\delta_1^1 + \delta_1^2 + \delta_{12} = 1, \quad \delta_1^1, \delta_1^2, \delta_{12} \geq 0$$

2. Player 2:

$$\max_{P_2^b,\delta_2^1,\delta_2^2,\delta_{21}} L \Rightarrow \max_{P_2^b,\delta_2^1,\delta_2^2,\delta_{21}} \underbrace{(p_1^a + p_2^a - p_1^b)}_{L_2^b} \tag{10.17}$$

subject to:

$$P_b^2(t) \leq P_{\max}$$

$$\delta_2^1 + \delta_2^2 + \delta_{21} = 1, \quad \delta_2^1, \delta_2^2, \delta_{21} \geq 0$$

Since the players do not communicate, they possess information only about their own decision variables. This gives rise to a game scenario. Since the decision variables are continuous, the power allocation problem is a continuous kernel non-zero sum game among the players. Since $(p_2^a - p_1^b - p_2^b)$ is a decreasing function of $P_a^1(t)$, the optimal value of $P_a^1(t)$ is $P_{\max}$. Using the same argument for the other players leads to the conclusion that the optimum level of power consumption of every player is $P_{\max}$. Therefore, the entire game terminates in a fixed time $T = \frac{E}{P_{\max}}$ irrespective of the initial positions of the agents.

Now we consider the problem of computing the optimal value of the decision variables for the players. In order to do so, we use pre-existing results from continuous kernel games that are presented here in Theorems 2 and 3 and are stated without proof.

Theorem 1 (Başar and Olsder 1999). *An N-person nonzero-sum game in which the finite-dimensional action spaces U^i $(i \in \mathbb{N})$ are compact and cost functionals J^i $(i \in \mathbb{N})$ are continuous on $U^1 \times \cdots \times U^N$ admits a mixed strategy Nash equilibrium (MSNE).*

From the above theorem, we can conclude that the power allocation game has a Nash equilibrium in mixed strategies since the decision variables of each player lie on a simplex, which is compact. Moreover, L is a continuous function of the decision variables of all the players. Therefore, the game admits a MSNE. Although, the MSNE has been computed for some games by exploiting some special characteristics in the cost functions, there are no standard techniques to compute MSNE for general continuous-kernel games (Owen 2001; Başar and Olsder 1999). Therefore, we search for the conditions under which the power allocation game admits a pure-strategy Nash equilibrium (PSNE).

Theorem 2 (Başar and Olsder 1999). *Let U be a closed, bounded and convex subset of $\mathbb{R}^m$, and for each $i \in \mathbb{N}$ the cost functional $J^i : U \to \mathbb{R}$ be continuous on U and convex in u^i for every $u^j \notin U^j$, $j \in \mathbb{N}, j \notin i$. Then, the associated N-person nonzero-sum game admits a PSNE.*

The above theorem provides the conditions under which we can guarantee existence of a PSNE. Let us consider the case of agent 1^a. The expressions for SINR provided in (10.13) relevant to the optimization problem being solved by 1^a can be written in a concise form as shown below:

$$s_2^a = a_1\gamma^{12}, \quad s_1^b = \frac{b_1}{c_1 + \gamma_1^1}, \quad s_2^b = \frac{d_1}{e_1 + \gamma_2^1},$$

where

$$a_1 = \frac{1}{\frac{\sigma}{P_{max}\rho(1+d^{12})^{-\alpha}} + \delta_1^2\left(\frac{1+d_1^2}{1+d^{12}}\right)^{-\alpha} + \delta_2^2\left(\frac{1+d_2^2}{1+d^{12}}\right)^{-\alpha}}, \quad b_1 = \delta_{21}\left(\frac{1+d_{12}}{1+d_1^1}\right)^{-\alpha},$$

$$c_1 = \frac{\sigma}{P_{max}\rho(1+d_1^1)^{-\alpha}} + \gamma_1^2\left(\frac{1+d_1^2}{1+d_1^1}\right)^{-\alpha}, \quad d_1 = \delta_{12}\left(\frac{1+d_{12}}{1+d_2^1}\right)^{-\alpha},$$

$$e_1 = \frac{\sigma}{P_{max}\rho(1+d_2^1)^{-\alpha}} + \gamma_2^2\left(\frac{1+d_2^2}{1+d_2^1}\right)^{-\alpha}.$$

Note that a_1, b_1, c_1, d_1 and e_1 are independent of the decision of 1^a.

Theorem 3. *The power allocation team game has a unique Nash equilibrium in pure strategies if the following conditions hold for Team A:*

$$g''(s_i^a) > 0, \tag{10.18}$$

$$g''(s_1^b) + \frac{2}{b_i}(c_i + \gamma_1^i)g'(s_1^b) < 0, \tag{10.19}$$

$$g''(s_2^b) + \frac{2}{d_i}(e_i + \gamma_2^i)g'(s_2^b) < 0, \tag{10.20}$$

and equivalent conditions hold for Team B:

$$g''(s_i^b) > 0, \tag{10.21}$$

$$g''(s_1^a) + \frac{2}{l_i}(m_i + \delta_i^1)g'(s_1^a) < 0, \tag{10.22}$$

$$g''(s_2^a) + \frac{2}{n_i}(o_i + \delta_i^2)g'(s_2^a) < 0, \tag{10.23}$$

where $i \in \{1, 2\}$. *The constants* b_2, c_2, d_2, e_i, l_i, m_i, n_i, *and* o_i *are obtained by re-writing the SINR expressions as done above; their expressions can be found in the Appendix.*

Proof. Let us consider the case of 1^a. From Theorem 2, we can conclude that a PSNE exists if L_1^a is convex in its arguments when the decision variables of the other team players are fixed. From Boyd and Vandenberghe (2004), L_1^a is convex if and only if $\nabla^2 L_1^a > 0$ (for Team B, L_i^b is concave if and only if $\nabla^2 L_i^b < 0$), where the Hessian $\nabla^2 L_1^a$ is given by (10.24).

$$\nabla^2 L_1^a = \begin{bmatrix} a_1^2 g''(s_2^a) & 0 & 0 \\ 0 & -\frac{b_1^2}{(c_1+\gamma_1^1)^4}[g''(s_1^b) + \frac{2}{b_1}(c_1 + \gamma_1^1)g'(s_1^b)] & 0 \\ 0 & 0 & -\frac{d_1^2}{(e_1+\gamma_2^1)^4}[g''(s_2^b) + \frac{2}{d_1}(e_1 + \gamma_2^1)g'(s_2^b)] \end{bmatrix} \tag{10.24}$$

This then says that $g''(s_2^a) > 0, g''(s_1^b) + \frac{2}{b_1}(c_1 + \gamma_1^1)g'(s_1^b) < 0$, and $g''(s_2^b) + \frac{2}{d_1}(e_1 + \gamma_2^1)g'(s_2^b) < 0$. The theorem then follows by following similar steps to verify $\nabla^2 L_2^a > 0$, $\nabla^2 L_1^b < 0$, and $\nabla^2 L_2^b < 0$. □

Applying the KKT conditions (Luenberger 1969), in addition to the assumptions provided in the theorem that guarantee strict convexity of L_1^a, gives us the following equations that need to be satisfied by the unique globally optimal solution ($\bar{\gamma}$):

$$\nabla L_1^a(\bar{\gamma}) + \sum_{i=1}^{3} \lambda_i \nabla h_i(\bar{\gamma}) + \eta \nabla h(\bar{\gamma}) = 0$$

$$\left.\begin{matrix} \lambda_i h_i(\bar{\gamma}) = 0 \\ \lambda_i, \eta \geq 0 \end{matrix}\right\} \quad i \in \{1, 2, 3\}$$

where

$$h_1(\bar{\gamma}) = -\gamma^{12} \leq 0, \quad h_2(\bar{\gamma}) = -\gamma_1^1 \leq 0, \quad h_3(\bar{\gamma}) = -\gamma_2^1 \leq 0,$$
$$h(\bar{\gamma}) = \gamma^{12} + \gamma_1^1 + \gamma_2^1 - 1 = 0$$

Now, we present the necessary and sufficient conditions for the solution to the optimization problem for the agents. Let us consider the case of 1^a. The assumptions in Theorem 3 regarding strict convexity of L_1^a render the KKT conditions to be necessary as well as sufficient for the unique global minimum.

To this end, we obtain:

$$\nabla L_1^a = \begin{bmatrix} a_1 g'(s_2^a) \\ \frac{b_1 g'(s_1^b)}{(c_1+\gamma_1^1)^2} \\ \frac{b_1 g'(s_2^b)}{(c_1+\gamma_2^1)^2} \end{bmatrix}, \nabla h(\bar{\gamma}) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \nabla h_1(\bar{\gamma}) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \nabla h_2(\bar{\gamma}) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \nabla h_3(\bar{\gamma}) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Since $\gamma \in \Delta^3$, at most three of the constraints can be active at any given point. Hence, the gradient of the constraints at any feasible point are always linearly independent.

If two of the three constraints among $\{h_1, h_2, h_3\}$ are active, then $\bar{\gamma}$ has a unique solution that is given by the vertex of the simplex that satisfies the two constraints. If only one of the constraints among $\{h_1, h_2, h_3\}$ is active, then we have the following cases depending on which constraint is active:

1. $h_1(\bar{\gamma}^1) = 0$: $\bar{\gamma}^1 = (0, \gamma_1^{1*}, 1 - \gamma_1^{1*})$ satisfies the following equations

$$g'(s_2^b)\frac{d_1}{[e_1 + (1 - \gamma_1^{1*})]^2} = g'(s_1^b)\frac{b_1}{[c_1 + \gamma_1^{1*}]^2} \tag{10.25}$$

2. $h_2(\bar{\gamma}^2) = 0$: $\bar{\gamma}^2 = (1 - \gamma_2^{1*}, 0, \gamma_2^{1*})$ satisfies the following equations

$$a_1 g'(s_2^a) = \frac{d_1 g'(s_2^b)}{(e_1 + \gamma_2^{1*})^2} \tag{10.26}$$

3. $h_3(\bar{\gamma}^3) = 0$: $\bar{\gamma}^3 = (1 - \gamma_1^{1*}, \gamma_1^{1*}, 0)$ satisfies the following equations

$$a_1 g'(s_2^a) = \frac{b_1 g'(s_1^b)}{(c_1 + \gamma_1^{1*})^2} \tag{10.27}$$

If none of the inequality constraints are active, then

$$\bar{\gamma}^4 = (\underbrace{1 - \gamma_1^{1*} - \gamma_2^{1*}}_{\gamma^{12*}}, \gamma_1^{1*}, \gamma_2^{1*}),$$

is the solution to the following equations:

$$a_1 g'(s_2^a) - \frac{b_1}{[c_1 + \gamma_1^{1*}]^2} g'(s_1^b) = 0, \quad a_1 g'(s_2^a) - \frac{d_1}{[e_1 + \gamma_2^{1*}]^2} g'(s_2^b) = 0 \tag{10.28}$$

Here, $\bar{\gamma}$ lies in the set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), \bar{\gamma}^1, \bar{\gamma}^2, \bar{\gamma}^3, \bar{\gamma}^4\}$.

An important point to note is that a_1, b_1, c_1, d_1 and e_1 depend on the decisions of the other players. Therefore, the computation of the decision variables depends on the value of the decision variables of the rest of the players. A possible way to deal with this problem is to use iterative schemes for computation of strategies.

Başar and Olsder (1999) provides some insights into the efficacy of such schemes from the point of view of convergence and stability. In this article, we assume that each agent has enough computational power so as to complete these iterations in a negligible amount of time compared to the total horizon of the game.

In the next section, we express the conditions for the existence of PSNE in terms of limitations imposed by the physical communications layer.

10.5 Existence of PSNE Under M-QAM Modulation Schemes

The bit error rate (BER) depends on the SINR, the modulation scheme, and the error control coding scheme utilized. Communications literature contains closed-form expressions and tight bounds that can be used to calculate $g(s)$ when the noise and interference are assumed to be Gaussian (Goldsmith 2005). For example, using uncoded M-QAM, where Gray encoding is used to map the bits into the symbols of the constellation, the BER can be approximated by Palomar et al. (2005)

$$g(s) \approx \frac{\zeta}{\log(M)} \mathcal{Q}\left(\sqrt{\beta s}\right), \tag{10.29}$$

where $\zeta = 4(1 - 1/\sqrt{M})$, $\beta = 3/(M-1)$, and $\mathcal{Q}(.)$ is the tail probability of the standard Gaussian distribution which can be expressed in terms of the error function erf:

$$\mathcal{Q}(x) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{x}{\sqrt{2}}\right).$$

The conditions of Theorem 3 depend, in this case, primarily on the employed modulation and coding schemes.

Theorem 4. *When all players employ uncoded M-QAM modulation schemes, the power allocation team game has a unique PSNE if the following condition is satisfied:*

$$\beta\rho P_{max}\left(\min\{d^{12}, d_{12}\}\right)^{-\alpha} < 3\sigma. \tag{10.30}$$

Proof. The conditions of Theorem 3 need to be satisfied for a unique pure-strategy solution to exist. We first verify those conditions for Player 1^a when uncoded M-QAM modulations are used. To this end, we differentiate (10.29) to obtain:

$$g'(s) = -\frac{\zeta\sqrt{\beta}}{2\log(M)\sqrt{2\pi s}}\exp\left(-\frac{\beta}{2}s\right), g''(s) = \frac{\zeta\sqrt{\beta}(1+\beta\sqrt{s^2})}{4\log(M)\sqrt{2\pi s^3}}\exp\left(-\frac{\beta}{2}s\right). \tag{10.31}$$

From (10.31), we conclude that condition (10.18) holds for any value of a_1 and γ^{12}. Condition (10.19) holds given that

$$c_1 > \frac{\beta}{3} b_1 - \gamma_1^1 \quad \Rightarrow \quad c_1 > \frac{\beta}{3} b_1, \tag{10.32}$$

which we can re-write as

$$\frac{\sigma}{P_{max}\rho} + \gamma_1^2 (d_1^2)^{-\alpha} > \frac{\beta}{3} \delta_{21} (d_{12})^{-\alpha}, \quad \frac{\sigma}{P_{max}\rho} > \frac{\beta}{3} (d_{12})^{-\alpha}. \tag{10.33}$$

By following similar steps, we can show that (10.33) is sufficient for (10.20) to hold. In fact, condition (10.33) is also sufficient for the convexity of L_2^a. For Team B, a sufficient condition for the concavity of L_1^b and L_2^b is

$$\frac{\sigma}{P_{max}\rho} > \frac{\beta}{3} (d^{12})^{-\alpha}, \tag{10.34}$$

which can be derived following similar steps to the above. The theorem follows from (10.33) and (10.34). □

Note that the left-hand side of inequality (10.30) depends entirely on physical design parameters; this is of particular importance for design purposes. Moreover, the sufficient condition of Theorem 4 can be expressed in terms of the received signal-to-noise ratios (SNRs) for all players, which could be more insightful from a communication systems perspective. Consider, for example, Player 1^a, and let $\text{SNR}_y^x = \frac{P_{max}\gamma_y^x \rho (d_y^x)^{-\alpha}}{\sigma}$ and $\text{SNR}_{xy} = \frac{P_{max}\delta_{xy}\rho (d_{yx})^{-\alpha}}{\sigma}$. Expression (10.32) can then be written as

$$\text{SNR}_{21} < \frac{3}{\beta} (\text{SNR}_1^2 + 1).$$

Similarly, condition (10.20) holds if

$$\text{SNR}_{12} < \frac{3}{\beta} (\text{SNR}_2^2 + 1).$$

Yet another useful way to interpret condition (10.30) is regarding it as a minimum rate condition:

$$R > \log \left(1 + \frac{\rho P_{max} \left(\min\{d^{12}, d_{12}\} \right)^{-\alpha}}{\sigma} \right),$$

where $R = \log(M)$.

The specific conditions for Player 1^a corresponding to (10.25)–(10.27) when M-QAM modulations are utilized are:

$$\left(\frac{s_1^b}{s_2^b}\right)^{\frac{3}{2}} \exp\left(-\frac{\beta}{2}(s_1^b - s_2^b)\right) - \frac{b_1}{d_1} = 0, \left(\frac{s_2^b}{s_2^a}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta}{2}(s_2^a - s_2^b)\right) - \frac{a_1 d_1}{(e_1 + \gamma_2^1)^2} = 0,$$

$$\left(\frac{s_1^b}{s_2^a}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta}{2}(s_2^a - s_1^b)\right) - \frac{a_1 b_1}{(c_1 + \gamma_1^1)^2} = 0.$$

Also, (10.28) in this case becomes

$$\left(\frac{s_2^b}{s_2^a}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta}{2}(s_2^a - s_2^b)\right) - \frac{a_1 d_!}{(e_1 + \gamma_2^1)^2} = 0,$$

$$\left(\frac{s_1^b}{s_2^a}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta}{2}(s_2^a - s_1^b)\right) - \frac{a_1 b_1}{(c_1 + \gamma_1^1)^2} = 0.$$

10.6 Conclusion and Future Work

This article has studied the power allocation problem for jamming teams. The motion of the teams was modeled using the framework of pursuit-evasion games and the optimal strategies were derived. An underlying static game was used to obtain the optimal power allocation, where the power budget of each user is split between communication and jamming activities. This work focused on the analysis of teams consisting of two players only, but as mentioned earlier at the conceptual level the analysis equally applies to the case when teams have more than two players. Still such an extension presents a plethora of interesting questions to address. In view of this, potential future directions include:

- *Computation of Singular Surfaces*: In this work, we have computed the trajectories based on the necessary conditions of optimality imposed by the Isaacs' conditions. In order to complete the construction of the optimal trajectories of the agents, we have to identify the singular surfaces in the state space (Başar and Olsder 1999). This is an interesting future research direction since the construction and nature of the singular surfaces would depend on the value of the decision variables obtained from the power allocation game.
- *Computation of MSNE*: As discussed in Sect. 10.3, the power allocation game admits a MSNE without any constraints on the underlying communication model. An important future problem is to compute the MSNE for the power allocation game.
- *Scheduling Schemes*: An interesting direction would be exploring scheduling algorithms, similar to the one proposed in Fu et al. (2010), in which players take turns in communicating or jamming. For example, the users of a given team

that are closest in distance to the other team could allocate all their resources to jamming, while the other users allocate all their resources to communicating with each other.

- *Power Control*: When a large number of users are present, and due to the broadcast nature of wireless systems, networks become interference-limited. The transmission power of one user can impede the links between other nodes due to the interference; hence, it is important to regulate the transmission power of the users in order to, for example, maximize the total capacity of the network.
- *Routing*: Multihop routing improves the total throughput and power efficiency of a network through relaying packets via intermediate nodes to their final destination. Because a portion of the energy of each node has to be allocated to jam the other team, determining the optimal route for transmission becomes a challenge, especially in the presence of mobility. An investigation of routing protocols in the context of games is therefore essential for studying the overall performance of the networks (Srivastava et al. 2005).
- *Eavesdropping:* When members of the two teams communicate at the same frequency, another security issue arises as the agents of a given team can receive and decode messages intended for internal communications of other teams. To ensure secure communications, each team would need to allocate power to jam the eavesdroppers. In fact, a more general scenario is when adversarial teams consist of active eavesdroppers: malicious nodes that can act as jammers and eavesdroppers (Mukherjee and Swindlehurst 2010).

Acknowledgements This research was supported in part by the U.S. Air Force Office of Scientific Research (AFOSR) MURI grant FA9550-10-1-0573, and Iowa State University research initiation grant.

Appendix

We include here the expressions for the constant parameters used in the statement of Theorem 3.

$$a_2 = \frac{1}{\frac{\sigma}{P_{max}\rho(1+d^{12})^{-\alpha}} + \delta_1^1\left(\frac{1+d_1^1}{1+d^{12}}\right)^{-\alpha} + \delta_2^1\left(\frac{1+d_2^1}{1+d^{12}}\right)^{-\alpha}}, \quad b_2 = \delta_{21}\left(\frac{1+d_{12}}{1+d_1^2}\right)^{-\alpha},$$

$$c_2 = \frac{\sigma}{P_{max}\rho(1+d_1^2)^{-\alpha}} + \gamma_1^1\left(\frac{1+d_1^1}{1+d_1^2}\right)^{-\alpha}, d_2 = \delta_{12}\left(\frac{1+d_{12}}{1+d_2^2}\right)^{-\alpha},$$

$$e_2 = \frac{\sigma}{P_{max}\rho(1+d_2^2)^{-\alpha}} + \gamma_2^1\left(\frac{1+d_2^1}{1+d_2^2}\right)^{-\alpha},$$

$$k_1 = \frac{1}{\frac{\sigma}{P_{max}\rho(1+d_{12})^{-\alpha}} + \gamma_2^1\left(\frac{1+d_2^1}{1+d_{12}}\right)^{-\alpha} + \gamma_2^2\left(\frac{1+d_2^2}{1+d_{12}}\right)^{-\alpha}}, \quad l_1 = \gamma^{21}\left(\frac{1+d^{12}}{1+d_1^1}\right)^{-\alpha}$$

$$m_1 = \frac{\sigma}{P_{max}\rho(1+d_1^1)^{-\alpha}} + \delta_2^1\left(\frac{1+d_2^1}{1+d_1^1}\right)^{-\alpha}, n_1 = \gamma^{12}\left(\frac{1+d^{12}}{1+d_1^2}\right)^{-\alpha},$$

$$o_1 = \frac{\sigma}{P_{max}\rho(1+d_1^2)^{-\alpha}} + \delta_2^2\left(\frac{1+d_2^2}{1+d_1^2}\right)^{-\alpha},$$

$$k_2 = \frac{1}{\frac{\sigma}{P_{max}\rho(1+d_{12})^{-\alpha}} + \gamma_1^1\left(\frac{1+d_1^1}{1+d_{12}}\right)^{-\alpha} + \gamma_1^2\left(\frac{1+d_1^2}{1+d_{12}}\right)^{-\alpha}}, l_2 = \gamma^{21}\left(\frac{1+d^{12}}{1+d_2^1}\right)^{-\alpha}$$

$$m_2 = \frac{\sigma}{P_{max}\rho(1+d_2^1)^{-\alpha}} + \delta_1^1\left(\frac{1+d_1^1}{1+d_2^1}\right)^{-\alpha}, n_2 = \gamma^{12}\left(\frac{1+d^{12}}{1+d_2^2}\right)^{-\alpha},$$

$$o_2 = \frac{\sigma}{P_{max}\rho(1+d_2^2)^{-\alpha}} + \delta_1^2\left(\frac{1+d_1^2}{1+d_2^2}\right)^{-\alpha},$$

References

Arnold VI (1983) Geometric method in the theory of ordinary differential equations. Springer, New York

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory, 2nd edn. SIAM series in classics in applied mathematics. SIAM, Philadelphia

Bhattacharya S, Başar T (2010a) Differential game-theoretic approach for spatial jamming attack in a UAV communication network. In: 14th international symposium on dynamic games and applications, Banff

Bhattacharya S, Başar T (2010b) Game-theoretic analysis of an aerial jamming attack on a UAV communication network. In: Proceedings of the American control conference, Baltimore, pp 818–823

Bhattacharya S, Başar T (2010c) Graph-theoretic approach to connectivity maintenance in mobile networks in the presence of a jammer. In: Proceedings of the IEEE conference on decision and control, Atlanta, pp 3560–3656

Bhattacharya S, Başar T (2010d) Optimal strategies to evade jamming in heterogeneous mobile networks. In: Proceedings of the workshop on search and pursuit-evasion, Anchorage

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge. Available at http://www.stanford.edu/~boyd/cvxbook/

Dong L, Han Z, Petropulu A, Poor H (2010) Improving wireless physical layer security via cooperating relays. IEEE Trans Signal Process 58(3):1875–1888

Fu L, Liew SC, Huang J (2010) Fast algorithms for joint power control and scheduling in wireless networks. IEEE Trans Wirel Commun 9(3):1186–1197

Goldsmith A (2005) Wireless communications. Cambridge University Press, Cambridge

Han Z, Marina N, Debbah M, Hjorungnes A (2009a) Physical layer security game: How to date a girl with her boyfriend on the same table. In: International conference on game theory for networks, 2009. GameNets '09, pp 287–294

Han Z, Marina N, Debbah M, Hjorungnes A (2009b) Physical layer security game: interaction between source, eavesdropper, and friendly jammer. EURASIP J Wirel Commun Netw (Special issue on Wireless Physical Layer Security) 2009(11):1–10

Isaacs R (1965) Differential games. Wiley, New York

Luenberger DG (1969) Optimization by vector space methods. Wiley, New York

Mukherjee A, Swindlehurst AL (2010) Equilibrium outcomes of dynamic games in mimo channels with active eavesdroppers. In: 2010 IEEE international conference on communications (ICC), pp 1–5

Owen G (2001) Game theory. Academic, London

Palomar DP, Bengtsson M, Ottersten B (2005) Minimum BER linear transceivers for mimo channels via primal decomposition. IEEE Trans Signal Process 53(8):2866–2882

Srivastava V, Neel J, Mackenzie A, Menon R, Dasilva L, Hicks J, Reed J, Gilles R (2005) Using game theory to analyze wireless ad hoc networks. IEEE Commun Surv Tutor 7(4):46–56

Tague P, Slater D, Noubir G, Poovendran R (2009) Linear programming models for jamming attacks on network traffic flows. In: Proceedings of 6th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt'08), Berlin

Chapter 11
Speculative Constraints on Oligopoly

Sébastien Mitraille and Henry Thille

Abstract We examine an infinite horizon game in which producers' output can be purchased by speculators for resale in a future period. The existence of speculators serves to constrain the feasible set of prices that can result from producers' output game in each period. Absent speculation, producers play a repeated Cournot game with random demand. With speculative inventories possible, the game becomes a dynamic one in which speculative stocks are a state variable which firms can control via their influence on price. We employ collocation methods to find the unknown expected price and value functions required for computation of equilibrium quantities. We demonstrate that strategic considerations result in an incentive to sell to speculators that is non-monotonic in the number of producers: speculation has the largest effect on equilibrium prices and welfare for market structures intermediate between monopoly and perfect competition. Using a computed example, we demonstrate that the effect of speculative storage on the average price level can be substantial, even though the effects on social welfare can be ambiguous.

Keywords Inventories • Oligopoly • Speculation • Dynamic game

MSC Codes: 91B24, 91B55, 91A25

S. Mitraille
Toulouse Business School, University of Toulouse, 20, Bd Lascrosses, Toulouse 31000, France
e-mail: s.mitraille@tbs-education.fr

H. Thille (✉)
Department of Economics & Finance, University of Guelph, 50 Stone Road East,
Guelph, ON, Canada N1G 2W1
e-mail: hthille@uoguelph.ca

© Springer International Publishing Switzerland 2016

F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*,
Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_11

11.1 Introduction

The impact of speculative storage on prices, profits and welfare, has recently received a surge of interest in the public debate, mostly due to substantial primary commodity price increases combined with the difficulty of consumers in developing countries to access some of these products. For example, the world oil market has been the object of recent political interest due to the sudden increase in speculation of the early 2000s. As pointed out in Smith (2009), this is an oligopolistic market dominated by OPEC with both "commercial" and "non-commercial" speculators active. Similarly, the impact of speculation on price, the importance of inventories, the access to important resources for developing countries, and the overall economic performance of commodity markets have been the subject of several recent debates. For example, the U.S. Senate committee on Homeland Security and Governmental Affairs pointed out in U.S. Senate (2006) that inventories of crude oil and natural gas have increased in the U.S. and in OECD countries due to an overall increase in speculation that sustained high prices and gave financial incentives to agents to store. According to this report, the inventory-price relationship has been perturbed compared to the usual negative correlation historically observed.[1] Likewise, the European Commission (2011) lists 14 critical raw materials[2] for which production is concentrated in the hands of few firms or a small number of countries. Finally, the formation of speculative bubbles on markets of vital or strategic importance for the development of emerging countries has attracted the attention of the United Nations Conference on Trade and Development (Gilbert 2010), for their crucial consequences on economic development and on the risks populations face. These questions have triggered substantial academic interest investigating the relationship between inventories and speculative trading on commodity markets from an econometric point of view (Frankel and Rose 2010; Kilian 2008, 2009; Kilian and Murphy 2014).

The effects of speculative storage when production is perfectly competitive is fairly well understood, with important contributions made in Newbery and Stiglitz (1981), Newbery (1984), Williams and Wright (1991), Deaton and Laroque (1992), Deaton and Laroque (1996), and McLaren (1999). The focus in these papers is on the effects of storage on the distribution of prices caused by the movement of production across periods due to random production (harvest) shocks. As aggregate inventories cannot be negative, speculators smooth prices across periods only when positive inventories exist. Unexpectedly large prices result in stock-outs which leads to a breakdown of the price smoothing role of speculative storage. These occasional stock-outs lead to a skewed distribution of price. Market power has been considered by examining imperfect competition in the storage function

[1] U.S. Senate (2006), p.15, Fig. 6.

[2] Antimony, Beryllium, Cobalt, Fluorspar, Gallium, Germanium, Graphite, Indium, Magnesium, Niobium, Platinum Group Metals, Rare earths, Tantalum, and Tungsten.

(Newbery 1984; Williams and Wright 1991; McLaren 1999), but production itself remains perfectly competitive in these papers, hence there is no scope for strategic considerations on the part of producers. While this approach is reasonable for modelling many agricultural commodities, where market power is often exhibited by intermediaries instead of primary producers, for many other commodities, such as the mineral and energy commodities discussed above, models with market power at the producer level are more appropriate.

The effects of speculation when there is imperfect competition at the producer level has been examined in Mitraille and Thille (2009) for monopoly production, and in Mitraille and Thille (2014) for oligopoly production,[3] although in a finite horizon setting. In Mitraille and Thille (2014) a two-period model of oligopoly production is used to demonstrate that speculative sales can result in a rich set of equilibria, including (1) stockouts, (2) deterrence of speculative holdings, (3) speculative holdings along with consumer purchases, (4) speculative purchases of the entire output, and (5) zero production.

Our contribution in this paper is to extend the analysis of Mitraille and Thille (2014) to an infinite horizon setting and to explore the implications of speculative storage on the price distribution under oligopolistic production. We do this by analyzing the Feedback equilibrium to an infinite horizon game in which oligopolists produce a commodity which can be purchased and stored for future sale by competitive speculators. We demonstrate that speculative storage can have significant effects on the distribution of prices and profits of an oligopoly compared to what would happen in the absence of storage. We find that for every market structure but monopoly, mean prices are lower or equal to the mean equilibrium price that occurs in the absence of speculative storage. Moreover the distribution of prices differences with and without speculative storage is asymmetric: prices below those of the Cournot equilibrium occur relatively frequently which means that speculative storage has a pro-competitive effect. When the number of firms increases, the price distribution converges to the Cournot one, but from below. Higher prices than those of the Cournot equilibrium are nonetheless possible: when the number of firms is low enough the equilibrium price may be high enough to exclude consumers from purchasing or to deter speculators from purchasing. This is particularly true when the market is monopolized, in which case the mean price is strictly higher when speculative storage is possible than when it is not.

We confirm these findings by studying the average profit deviation from Cournot competition absent competitive speculation: profit is the smallest compared to Cournot when the number of firms is intermediate, while profits converge to Cournot from below when the number of firms increases. Despite the gains to an oligopoly due to price and cost smoothing, the presence of competitive speculators increases competition and lowers profits compared to Cournot. Similar results can be found

[3]The effects of producer storage on the equilibrium in a Cournot duopoly is examined in Thille (2006), in which, rather than speculators engaging in storage, producers themselves store in the face of random variations in demand and cost.

when comparing consumers surplus and total welfare to Cournot competition: the average gain in consumers surplus is the largest for an intermediate market structure.

In what follows we first describe the model and then explore the implications of speculative storage for the nature of equilibria that we expect to find. We then describe the computational approach that we take to finding the Feedback equilibrium to the game and finish with a description of the equilibrium for a computed example.

11.2 The Model

The model that we present here is an infinite horizon version of that in Mitraille and Thille (2014). We consider a discrete time model with an infinite number of periods in which risk-neutral consumers, producers, and speculators interact on the market for a homogeneous, non-perishable product. We assume that consumers and speculators are price takers and behave competitively, while a finite number n of producers behave as an oligopoly. Speculators are able to store the product while producers and consumers cannot.

In every period t, consumers have a demand, D_t, which they can buy on a spot market. Consumers' demand in period t, D_t, is a decreasing function of the spot price p_t, and is an increasing function of a random state a_t. We assume that consumer's demand is a linear function of p_t and a_t, given by

$$D(a_t, p_t) = \max\{a_t - p_t, 0\} \tag{11.1}$$

where the random state a_t is drawn by Nature at the beginning of period t and known to every participant of the spot market before decisions are made. We assume that the random states $\{a_t\}$ are independently and identically drawn from period to period as the random variable $\tilde{a}$, distributed over the support $[0, A]$ with a continuous cumulative distribution function $F(a)$, with $f(a)$ the associated density function.[4] We denote the mean of $\tilde{a}$ by

$$E(a) = \int_0^A a\, dF(a). \tag{11.2}$$

Random changes in a_t may be interpreted as random shocks affecting the distribution of income in the population of consumers from period to period, modifying in turn the willingness to pay for the product sold by firms and stored by speculators.

In every period t, speculators are able to buy or sell on the spot market, and are able to store the product. Let x_t denote the position of speculators on the spot market of period t: if x_t is positive, then speculators are selling the product, while

[4] In Mitraille and Thille (2014) a uniform demand is considered.

if x_t is negative speculators are buying the product. Speculators are able to store the product and we denote by S_t the amount of available inventories at the beginning of period t. This amount S_t is observable to all market participants. We assume that the rate of depreciation of inventories is constant and equal to γ; the transition equation for inventories is then

$$S_{t+1} = (1-\gamma)\,(S_t - x_t). \tag{11.3}$$

Negative inventories are not allowed and initial inventories are equal to 0, $S_0 = 0$. Consequently in every period t aggregate speculative sales must satisfy

$$x_t \in (-\infty, S_t]. \tag{11.4}$$

We assume that the cost of storage of speculators paid in every period is a linear function of the level of initial inventories held in that period, and equal to

$$W(S_t) = wS_t \tag{11.5}$$

with $w \geq 0$. Let the discount factor be $\delta \in (0,1]$, and let E_t denote the expectation operator conditional on the information available in period t.

In every period t, there are n producers in Cournot competition, each of which chooses the quantity it wants to produce, $q_t^i \in \mathbb{R}^+$, $i = 1, \ldots, n$. All firms produce their output using the same technology which results in the cost function

$$C(q_t^i) = \frac{c}{2}(q_t^i)^2 \tag{11.6}$$

with $c > 0$.

Firms cannot store their production: the quantity they produce in any period is equal to the quantity they sell on the market. We denote the aggregate quantity produced in period t by Q_t, and the aggregate quantity produced by all firms but i by Q_t^{-i}, where $Q_t = \sum_{i=1}^n q_t^i$ and $Q_t^{-i} = \sum_{j=1, j\neq i}^n q_t^j$. The vector of individual producer outputs will be denoted $q_t = (q_t^1, q_t^2, \ldots, q_t^n)$. Let p_t denote the market price, then producer i's instantaneous profit in period t is equal to

$$\pi_t^i = p_t q_t^i - C(q_t^i) \tag{11.7}$$

and the total expected profit discounted in period 0, Π_0^i, is

$$\Pi_0^i = E_0 \sum_{t=0}^{\infty} \delta^t \pi_t^i \;\text{ for all } i = 1, \ldots, n, \tag{11.8}$$

where E_t denotes the expectation operator conditional to the information available to all agents in period t.

The timing of the game adapts the Cournot timing to our dynamic setting where long-lived speculators have rational expectations over future prices. We assume that speculative inventories, S_t, and the demand state, a_t, are observed by all agents at the beginning of period t. Consequently, information is symmetric across agents. In period t, the timing of the interaction is therefore:

1. Demand level, a_t, is realized and observed by all agents. Aggregate inventory holdings, S_t, is observed by all agents.
2. Producers choose q_t^i, $i = 1, 2, \ldots, n$.
3. Speculators choose x_t.
4. Auctioneer sets p_t such that $D_t - x_t = q_t$.
5. Transactions occur and stage payoffs are realized.

Finally we assume that producers play stationary Feedback, or Markov, strategies,[5] depending only on the current state, (a_t, S_t). Producer i's strategy, σ^i, is a mapping from the set of states (a_t, S_t) to the set of quantities, $\sigma^i : [0, \overline{A}] \times \mathbb{R}^+ \to \mathbb{R}^+$. Given a strategy for each producer, $\sigma \equiv (\sigma^1, \ldots, \sigma^n)$, define $V^i(\sigma) = E_0 \sum_{t=0}^{+\infty} \delta^t \pi^i(\sigma, a_t, S_t)$ be the payoff to producer i under the strategy profile σ. Then,

Definition 1. A Feedback equilibrium with rational expectations is a n-tuple of strategies $\sigma^* \equiv (\sigma^{1*}, \ldots, \sigma^{n*})$ such that

$$V^i(\sigma^*) \geq V^i((\sigma^{1*}, \ldots, \sigma^{(i-1)*}, \sigma^i, \sigma^{(i+1)*}, \ldots, \sigma^{n*})) \ \forall\ \sigma^i \text{ for all } i = 1, \ldots, n \tag{11.9}$$

with, for every period t, inventories in period $t+1$ follow

$$S_{t+1}^* = (1-\gamma)\left(S_t^* - X^*(\sum \sigma^i(a_t, S_t^*), a_t, S_t^*)\right), \tag{11.10}$$

the market price p_t^* clears the market:

$$D_t^* - X^*(\sum \sigma^i(a_t, S_t^*), a_t, S_t^*) = \sum \sigma^i(a_t, S_t^*), \tag{11.11}$$

and the future market price $E_t[p_{t+1}^*]$ is rationally expected by all agents.

In a Feedback equilibrium, conditioning strategies to past prices or quantities is ruled out, so strategies allowing firms to implement tacit collusion are not considered.

[5] See Başar and Olsder (1995).

11.3 Speculators' Behaviour and Firms' Strategies

As in standard commodity storage models, speculators' behaviour is driven by the relationship between current and expected future prices. Speculators maximize their profit taking the current price as given and expecting the future price that results from the quantity of inventory carried into the next period, S_{t+1}. It will be useful to introduce the notation $p^e(S_{t+1}) = E_t[p_{t+1}|S_{t+1}]$ to represent the expected future price conditional on the level of stocks carried into $t+1$. As the behaviour of speculators determines the demand that will be faced by producers, we need to determine speculative sales as a function of producers' output. As the derivation is the same as for the finite horizon case, here we present a brief description of it. For a more detailed derivation see Mitraille and Thille (2014).

The aggregate behaviour of these speculators ensures that $p_t \geq \delta(1-\gamma)(p^e(S_{t+1})-w)$ with a stock-out occurring if the inequality is strict. The non-negativity constraint on aggregate speculative inventories implies that speculators' aggregate behaviour satisfies the complementarity condition

$$(S_t - x_t)(p_t - \delta(1-\gamma)(p^e(S_{t+1})-w)) = 0,$$

$$S_t - x_t \geq 0, \quad p_t - \delta(1-\gamma)(p^e(S_{t+1})-w) \geq 0 \tag{11.12}$$

Either no inventories are carried ($S_{t+1} = 0$) and the return to storage is negative, or inventories are carried ($S_{t+1} > 0$) and the return to storage is zero. Using $X^*(Q_t, a_t, S_t)$ to denote the equilibrium storage undertaken when producers sell Q_t in aggregate and the state is (a_t, S_t), the market clearing price, $P(Q_t, a_t, S_t)$, must be such that the total of consumer and speculative purchases satisfy

$$a_t - P(Q_t, a_t, S_t) - X^*(Q_t, a_t, S_t) = Q_t. \tag{11.13}$$

From (11.12), there is a threshold level of aggregate output, which we denote Q_t^L, below which $p_t > \delta(1-\gamma)(p_t^e(0)-w)$, as speculators cannot carry negative inventories. This threshold is the level of output which leads to zero return to speculation when there is a stockout:

$$a_t - S_t - Q_t^L = \delta(1-\gamma)(p_t^e(0)-w). \tag{11.14}$$

It is also possible that speculators purchase the entire production of firms, resulting in zero consumer purchases. This exclusion of consumers will occur if speculators value producers' output more highly than consumers do: $\delta(1-\gamma)(p_t^e(S_t+Q_t)-w) > a_t$. Consequently, there is another threshold output, which we denote $\widehat{Q}_t$, for which only speculators buy if $Q_t < \widehat{Q}_t$ and consumers buy and speculators carry inventories if $Q_t > \widehat{Q}_t$. This threshold is determined by the level of aggregate output that just extinguishes consumer demand when that output is purchased entirely by speculators:

$$a_t = \delta(1-\gamma)(p_t^e(S_t+\widehat{Q}_t)-w). \tag{11.15}$$

As with Q_t^L, the slope of the demand faced by producers changes discontinuously at $\widehat{Q}_t$. It is important to note that only one of Q_t^L and $\widehat{Q}_t$ can be positive as long as $p_t^e()$ is a decreasing function.[6]

Finally, when aggregate output exceeds the relevant threshold, Q_t^L or $\widehat{Q}_t$, speculative sales are determined implicitly by the relationship between p_t and $p_t^e(S_{t+1})$ that must hold. We denote speculative sales in this case as $\widetilde{X}(Q_t, a_t, S_t)$, which is the solution in X to

$$a_t - X - Q_t = \delta(1-\gamma)\left(p_t^e(S_t - X) - w\right). \tag{11.16}$$

Summarizing, speculative sales are given by

$$X^*(Q_t, a_t, S_t) = \begin{cases} S_t & \text{if} \quad Q_t \leq Q_t^L \text{ and } Q_t^L > 0 \\ -Q_t & \text{if} \quad Q_t \leq \widehat{Q}_t \text{ and } \widehat{Q}_t > 0 \\ \widetilde{X}(Q_t, a_t, S_t) & \text{otherwise.} \end{cases} \tag{11.17}$$

As only one of Q_t^L and $\widehat{Q}_t$ can be positive, only one of the first two conditions in (11.17) is possible for a given (S_t, a_t).

Given the behaviour of speculators in (11.17), we can now state the inverse demand function faced by producers:

$$P(Q_t, a_t, S_t) = \begin{cases} a_t - S_t - Q_t & \text{if} \quad Q_t \leq Q_t^L \text{ and } Q_t^L > 0 \\ \delta(1-\gamma)(p_t^e(S_t + Q_t) - w) & \text{if} \quad Q_t \leq \widehat{Q}_t \text{ and } \widehat{Q}_t > 0 \\ a_t - \widetilde{X}(Q_t, a_t, S_t) - Q_t & \text{otherwise.} \end{cases} \tag{11.18}$$

We plot this inverse demand for alternative demand levels in Figs. 11.1 and 11.2. Figure 11.1 illustrates a situation with $Q_t^L > 0$, in which speculators sell their entire stock of inventories when aggregate production is low, shifting down demand in a parallel fashion. Once aggregate production exceeds Q_t^L, a stockout no longer occurs and price is linked to the expected future price, which is less steeply sloped than consumer demand. Figure 11.2 illustrates demand for the same level of speculative inventories but with a lower demand state. Here, for low levels of aggregate production consumers do not buy any output as price exceeds their maximal willingness to pay of 2.5 and the entire production is purchased and stored by speculators. Once aggregate production exceeds $\widehat{Q}_t$ price is low enough to induce consumer purchases on top of speculators demand.

With the behaviour of speculators determined by (11.17) and (11.18), payoffs to producers can be specified entirely in terms of output. The marginal payoff to a firm

[6] If both thresholds were positive, both (11.14) and (11.15) must hold. The left hand side of (11.15) is clearly higher than that of (11.14) while the right hand side of (11.15) is lower than that of (11.14) if the expected price function is decreasing in future stocks, so both (11.14) and (11.15) cannot hold simultaneously.

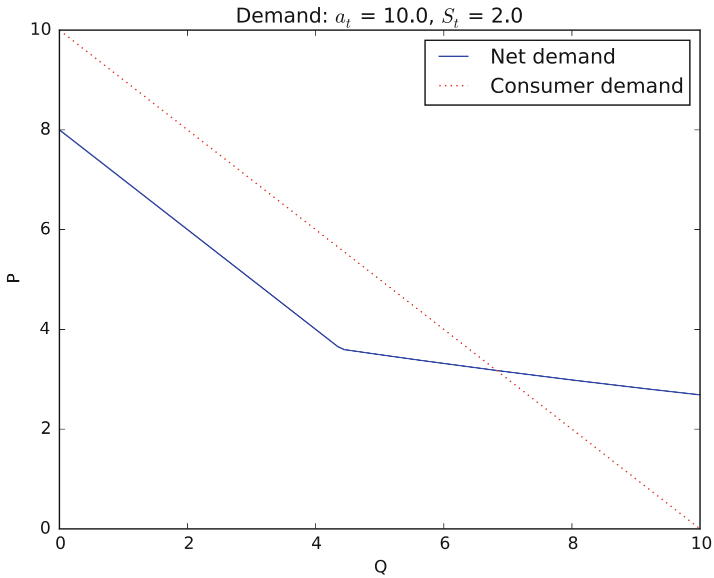

Fig. 11.1 Net demand and consumer demand when $Q_t^L > 0$. Net demand is generated using the same parameters as in the numerical solution reported below with the number of firms set at three

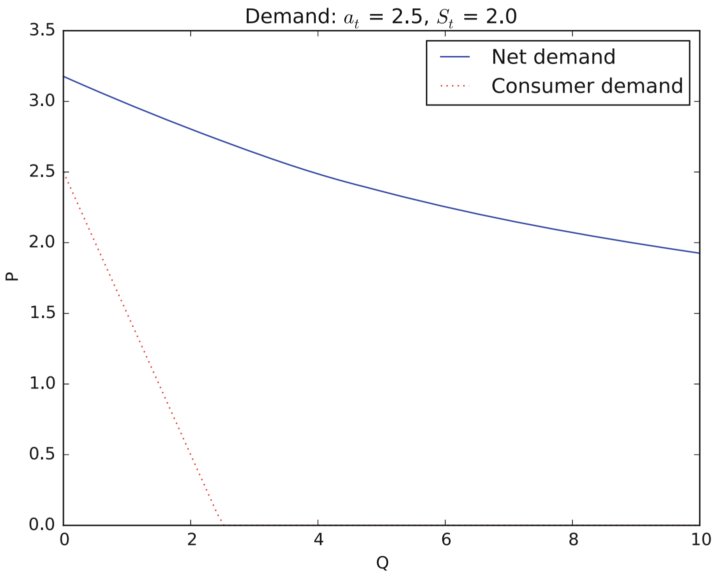

Fig. 11.2 Net demand and consumer demand when $\hat{Q}_t > 0$. Net demand is generated using the same parameters as in the numerical solution reported below with the number of firms set at three

in any period will be discontinuous at an output that results in aggregate output of Q_t^L or $\widehat{Q}_t$, so the nature of the equilibrium in period t depends on where aggregate output falls in relation to these thresholds. First, a stockout may occur in which speculators sell their entire stocks and consumers buy the total of speculative and producer sales. We will denote this type of equilibrium with a C. Second, consumers may buy nothing and speculators purchase the entire output of firms, which we will denote with an S. Third, consumers may make some purchases and speculators carry inventory into the following period, denoted with a CS. Finally, there is the possibility that producers deter speculators by producing exactly Q_t^L in aggregate, which we will denote with an L.

With the effects of speculation on the demand faced by producers determined, we can write the profit received by a producer in period t as

$$\pi^i(q_t, a_t, S_t) = P(Q_t, a_t, S_t)q_t^i - \frac{c}{2}(q_t^i)^2 \tag{11.19}$$

and the producer's payoff as

$$\Pi_0^i = E_0 \sum_{t=0}^{+\infty} \delta^t \pi^i(q_t, a_t, S_t) \text{ for } i = 1, \ldots, n. \tag{11.20}$$

Given the behaviour of speculators determined above, we can express the dynamics of speculative inventory as depending on production choices:

$$S_{t+1} = (1-\gamma)(S_t - X^*(Q_t, a_t, S_t)). \tag{11.21}$$

Consequently, the game played by producers has payoffs given by (11.20) and state dynamics given by (11.21).

As in Mitraille and Thille (2014), producer payoffs, while continuous, are non-differentiable at the thresholds Q_t^L and $\widehat{Q}_t$. As proven in that paper, in the context of the two-period game starting in period T-1, the fact that payoffs are not differentiable at threshold output levels Q_t^L and $\widehat{Q}_t$ generate discontinuities in the marginal profits, which results in upward jumps at $\widehat{Q}_t$, as well as upward or downward jumps at Q_t^L. This implies that profit comparisons must be performed in order to determine which of the different potential equilibria exist. For example when $\widehat{Q}_t > 0$, the equilibrium can be either the one with consumer exclusion (S), or the one with consumer and speculative purchases (CS), and firms profits must be compared to determine which one occurs, with potentially a multiplicity of outcomes when none of the local equilibrium candidates can be ruled out by a global unilateral deviation. Similarly, when $Q_t^L > 0$, profit comparisons must be performed to determine which of the candidates, between a stock-out equilibrium (C) and an equilibrium with consumers and speculative purchases (CS), is the equilibrium to the game. Moreover, in this case, the discontinuities in marginal profits also implies that an equilibrium with aggregate output equal to Q_t^L may exist. These forces, demonstrated in Mitraille and Thille (2014) for the two period game, still exist in the infinite horizon game. The determination of the equilibrium of the game for a given set of parameters requires verification of which of the potential candidates, C, S, CS, or L, exist.

In order to solve the game, we need to determine the expected price function and the value function associated with the equilibrium production strategies of producers. The expected price function is required to determine the thresholds Q_t^L and $\widehat{Q}_t$ as well as speculative sales when prices are smoothed [from (11.16)] and is computed as

$$p_t^e(S_{t+1}) = \int_0^A p^*(a, S_{t+1})dF(a) \tag{11.22}$$

where $p^*(a, S)$ is the Feedback equilibrium price when producers play equilibrium strategies $\sigma^*(a, S)$. The value function associates the expected payoff to a firm under the equilibrium strategies when the current state is (a_t, S_t):

$$V(a_t, S_t) = p^*(a_t, S_t)\sigma^*(a_t, S_t) - \frac{c}{2}\sigma^*(a_t, S_t)^2 + \delta E_t[V(a_{t+1}, S_{t+1})], \tag{11.23}$$

with $S_{t+1} = (1 - \gamma)(S_t - X^*(n\sigma^*(a_t, S_t), a_t, S_t))$. We next turn to describing the method we use to compute these functions.

11.4 Numerical Approach

Following the strategy used by Williams and Wright (1991), who compute approximations to the smooth $p^e(S_{t+1})$ rather than $p^*(a_t, S_t)$ for the competitive case, we approximate $p^e(S_{t+1})$ as well as the expected value function:

$$V^e(S_{t+1}) = \int_0^A V(a, S_{t+1})dF(a). \tag{11.24}$$

We apply the collocation method,[7] using cubic splines to approximate the expected price and value functions, denoting these approximations ρ and ν. We start with the period T solution[8] from Mitraille and Thille (2014) and iterate until convergence of the expected price and value function approximations. Given a vector of m values

[7]Judd (1998), Chap. 11 provides a description of the method, which he applies to the competitive storage problem in Chap. 17.4.

[8]In order to facilitate convergence of the expected value function we replace the terminal value of zero in Mitraille and Thille (2014) with the value of an infinite stream of Cournot profit following some "terminal time" beyond which speculation is not possible. In consequence, the initial condition for the value function is that which occurs in a period in which speculators are forced to sell their inventory and unable to replenish it again.

for S_{t+1}, $\bar{S} \equiv (0, S_1, \ldots, S_{max})$, in the jth iteration we compute the quantities $\bar{p}^e_{jk}$ and $\bar{V}^e_{jk}$ which are the expected price and value associated with future stocks given by each element of $\bar{S}$, i.e., for $k = 1, 2, \ldots, m$,

$$\bar{p}^e_{jk} = \int_0^A p^*_{j-1}(a, \bar{S}_k) dF(a) \tag{11.25}$$

and

$$\bar{V}^e_{tk} = \int_0^A V_{j-1}(a, \bar{S}_k) dF(a) \tag{11.26}$$

with $p^*_{j-1}(a, S)$ and $V_{j-1}(a, S)$ the equilibrium price and value functions found in iteration $j-1$ using the iteration $j-1$ approximations ρ_{j-1} and ν_{j-1}. Our approximation to $p^e_j(S)$, $\rho_j(S)$, is found by fitting a cubic spline to the $\bar{S}$ and $\bar{p}^e_j$ vectors. Similarly, we fit a cubic spline to $\bar{S}$ and $\bar{V}^e_j$ to generate our approximation to V^e_j, $\nu_j(S)$.

In summary, to find the solution in any period t_0:

Step 0 Compute the iteration 0 equilibrium price and value, $p^*_0(a, \bar{S}_k)$ and $V_0(a, \bar{S}_k)$, $k = 1, \ldots m$, using (11.22) and (11.23). Set $j = 1$.
Step 1 For each $k = 1, \ldots, m$ compute

$$\bar{p}^e_{jk} = \int_0^A p^*_{j-1}(a, \bar{S}_k | \rho_{j-1}, \nu_{j-1}) dF(a), \tag{11.27}$$

$$\bar{V}^e_{jk} = \int_0^A V^*_{j-1}(a, \bar{S}_k | \rho_{j-1}, \nu_{j-1}) dF(a) \tag{11.28}$$

Step 2 Fit a cubic spline to $(\bar{S}, \bar{p}^e_j)$ and $(\bar{S}, \bar{V}^e_j)$ to form $\rho_j(S)$ and $\nu_j(S)$.
Step 3 Return to Step 1 until $\rho_j(S)$ and $\nu_j(S)$ have not changed appreciably from the previous iteration.

When computing the equilibrium for any iteration an equilibrium selection is required for the cases in which multiple equilibria occur. We assume producers play C when both C and CS are possible and CS when both CS and S are possible.[9]

11.5 Results

We solve the infinite horizon game for the same values of the model parameters used in Mitraille and Thille (2014), namely $\delta = 0.95$, $w = 0.2$, $\gamma = 0$, and $c = 0.6$. The random demand parameter is distributed uniform on [0, 20]. For the cubic spline interpolations used in approximating the expected price and value functions, we use

[9]The code used to compute the solution uses numerical routines from NumPy and SciPy (Jones et al. 2001). The code is available from the authors on request.

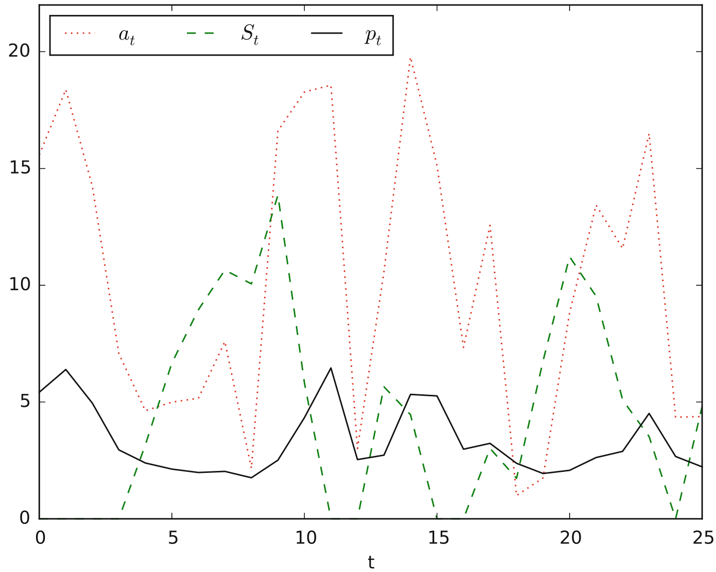

Fig. 11.3 Sample time series plot of the demand state, a_t, beginning of period stocks, S_t, and price, p_t, for the case $n = 3$

a grid of 25 values for the level of speculative stocks on a range between 0 and 60. We present results from the solution obtained after 50 iterations, by which time the maximal change in the expected value function is of the order of 0.1 %.[10]

In order to demonstrate the effects of speculation on oligopoly, we present statistics gathered from running simulations of 1000 periods for each n and computing statistics of interest. A short sample of a simulated time series with three producers is presented in Fig. 11.3 in which we see instances of the alternative equilibria. Large realizations of a_t are often associated with relatively high price and a stockout. For example, periods 11 and 15 are ones in which the C equilibrium occurs. It is interesting to note that a large a_t is not sufficient to generate a stockout: in periods 9 and 10 there are relatively high realizations of a_t but a stockout has not occurred due to the rather high level of inventory that was built up during a sequence of below average realizations of a_t in periods 3–8. We also see examples of zero consumer purchases in Fig. 11.3. For example, periods 18 and 19 see $p_t > a_t$ which implies zero consumption. Producers are selling solely to speculators at this time resulting in a rapid accumulation of speculative stocks. The overall frequencies of the alternative equilibria from this simulation with three firms are 26.2 % C, 64.1 % CS, and 9.7 % S, with no occurrences of the L equilibrium.[11]

---

[10]The expected price function converges much more quickly than the expected value function.

[11]It is demonstrated in Mitraille and Thille (2014) that the L equilibrium is unlikely to occur when there are few, but more than one, firms.

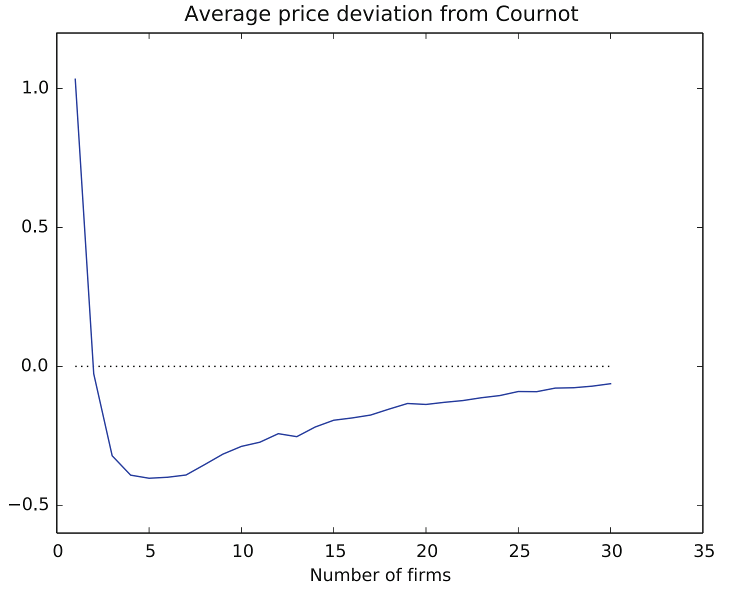

Fig. 11.4 Average difference between price with speculators and that without speculators over a 1000 period simulation

We now summarize the effects of speculation by examining deviations of the quantities of interest in the game with speculation from that which occurs in the absence of speculation (the repeated Cournot game with random demand[12]). First, in order to see what effect speculators have on price levels, we plot the average deviation from the Cournot price in Fig. 11.4. Consistent with Mitraille and Thille (2009), speculation has an anti-competitive effect on prices in the case of monopoly. The increase in average price for $n = 1$ in Fig. 11.4 is roughly 17 % of the mean price in the absence of speculation. In this case, a monopolist's desire to price in such as way as to limit the building up of speculative inventories results in substantially higher prices on average. However, for more than one firm, speculation has a pro-competitive effect. In the oligopoly setting, firms compete to sell to speculators (effectively competing to supply future demand) resulting in prices lower than in the absence of speculation. Again, this effect is not trivial, the gap between prices

[12]In using this benchmark, we are examining the effects of adding speculators with a storage technology to a model in which no storage technology exists. To examine the effects of speculation alone, we would need to allow producers to store the good, which would be a substantial complication. However, Thille (2006) has shown that, in the absence of speculators, the average price level was the same in the equilibrium in which producers can store the commodity as in the equilibrium in which they could not store. Consequently, we are confident that the results we present below are predominantly due to the presence of speculation and not simply due to the addition of a storage technology.

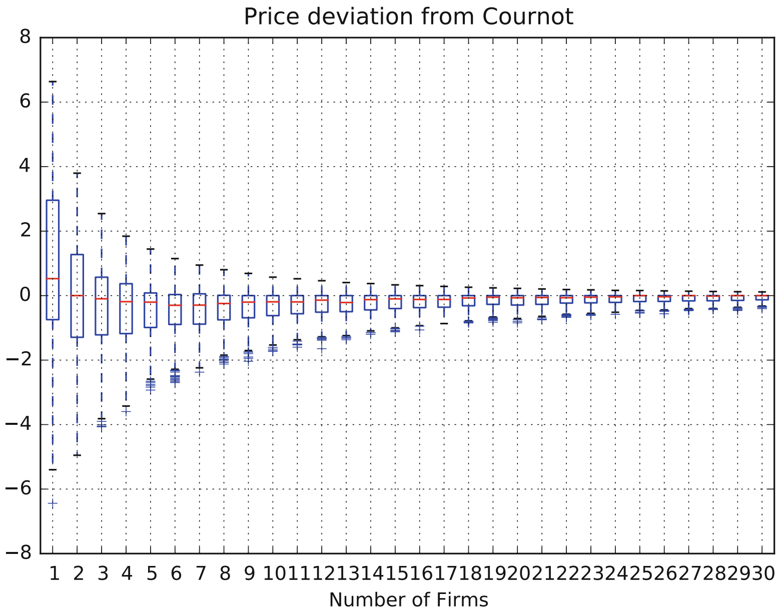

Fig. 11.5 Box and whisker plot for the difference in price between the model with and without speculators. The box extends from the lower to the upper quartile, while the whiskers are set at 1.5 times the inter-quartile range

with and without speculation in Fig. 11.4 is more than 20 % in some cases. Hence, simply by decoupling sales to consumers from production over time, speculation has a substantial effect on the average level of price in an oligopoly. It is important to note that in a similar setting in which there are no speculators, but producers themselves have the ability to store, Thille (2006) demonstrates that the average level of price is not affected by the addition of a storage technology for producers. Consequently, we attribute the effects on average price in Fig. 11.4 to the presence of speculators with a storage technology, and not to the addition of the storage technology alone.

In order to illustrate the distribution of the deviation of prices from the Cournot equilibrium, box-plots of price are plotted for each n in Fig. 11.5. Price deviations from Cournot tend to be asymmetrically distributed, negative values being more frequent than positive ones for $n > 1$. The opposite occurs when production is monopolized, due to the fact that a monopoly selects more often the limit equilibrium compared to a more competitive market structure. As the number of firms in competition increases, the price distribution converges to that of the unique Cournot equilibrium price, but largely from below. Figure 11.5 also illustrates that the magnitude of the effect of speculation on price can be quite large for relatively concentrated market structures.

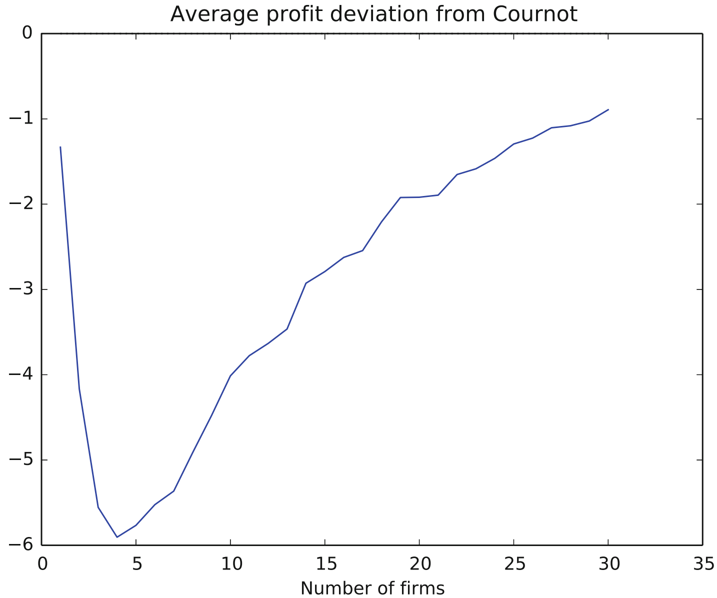

Fig. 11.6 Average difference between profit with speculators and that without speculators over a 1000 period simulation

The effects of speculation on mean profits are presented in Fig. 11.6. Not surprisingly, the pro-competitive effect of speculation that results in lower average prices in an oligopoly also reduces producer profits relative to the case in which speculation is absent. This is true even for monopoly. By smoothing prices over time, speculators restrict to some extent the ability of the monopolist to realize maximal profit. This is easiest to see in the limit equilibrium, where the monopolist chooses a level of output that is lower than the one that maximizes profit in the absence of speculation in order to deter speculative purchases. Although the monopolist does limit speculation this way, it still earns lower profit than it would if speculators were absent.

Given these non-monotonic effects of speculation on price and profit, it is interesting to examine the net effect on consumer surplus and welfare. These are plotted in Figs. 11.7 and 11.8. The average consumer surplus, depicted in Fig. 11.7, is positive for any number of firms. The gain is relatively low for $n = 1$, rises to a maximum at $n = 3$, and then declines slowly as the number of firms increases. It is interesting to note that for $n = 1$, consumers benefit from speculation even though average price is higher. This is due to the variation in the effects of speculation as shown in Fig. 11.5 and the fact that price affects consumer surplus in a non-linear manner. Intuitively, in periods of high demand (large a_t) prices tend to be higher, causing speculators to sell their stocks. Hence, speculators dampen price when it has the largest effect on consumer surplus. The limit equilibrium does not exist at high levels of demand, so the situations in which price is increased due to speculation

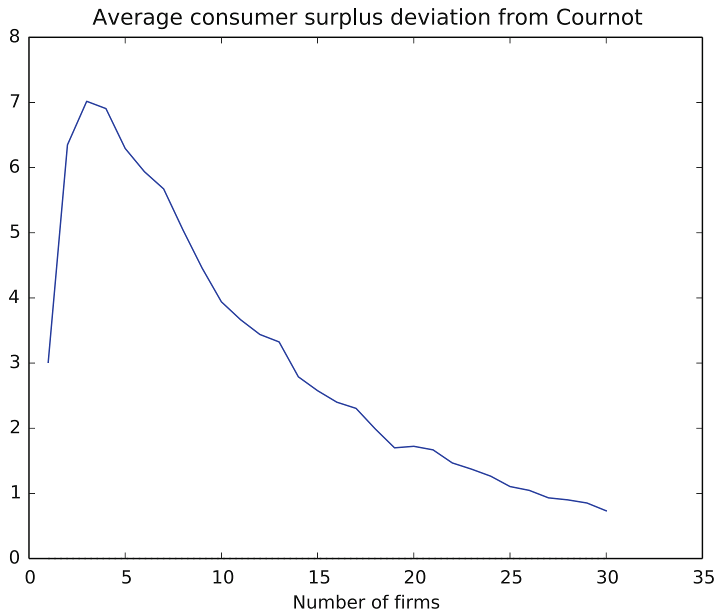

Fig. 11.7 Average difference between consumer surplus with speculators and that without speculators over a 1000 period simulation

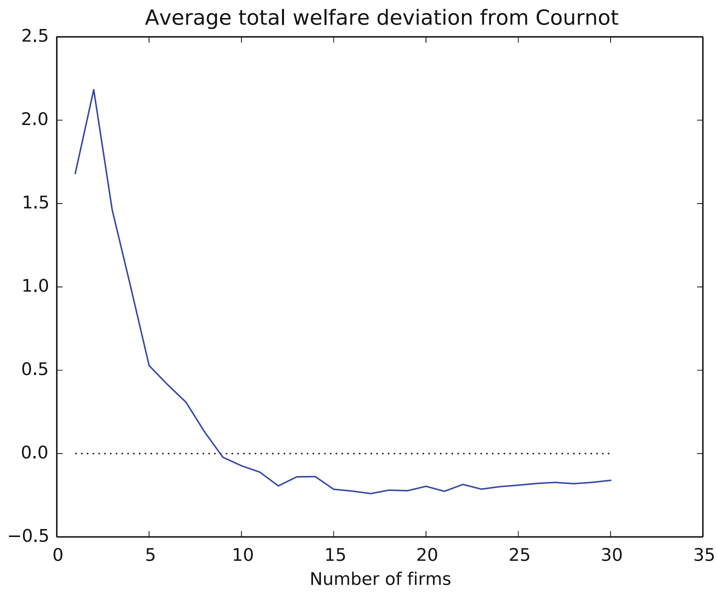

Fig. 11.8 Average difference between welfare with speculators and that without speculators over a 1000 period simulation

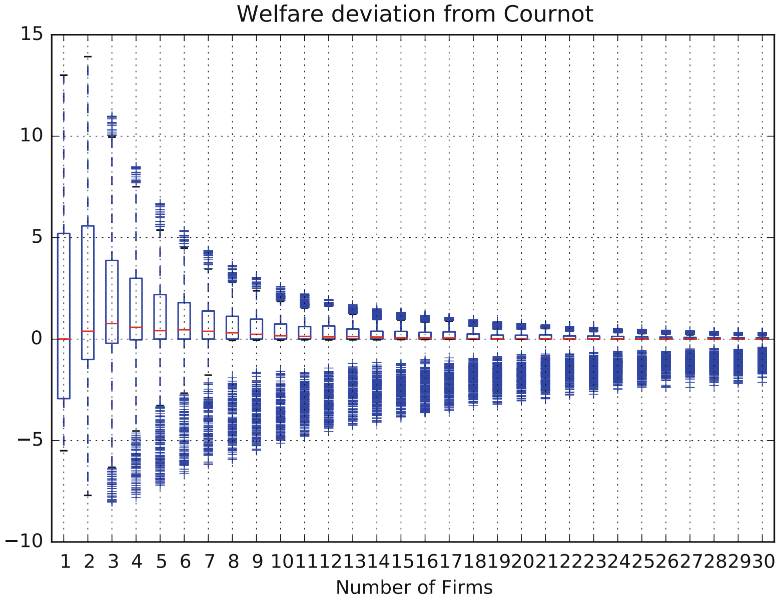

Fig. 11.9 Box and whisker plot for the difference in welfare between the model with and without speculators. The box extends from the lower to the upper quartile, while the whiskers are set at 1.5 times the inter-quartile range

occur in states where demand is lower, having a smaller effect on consumer surplus. For $n > 1$ this effect is complemented by the average reduction in price due to speculation.

Combining the effects of speculation for both consumers and producers, we see from Fig. 11.8 that the average effect of speculation for social welfare is ambiguous. For relatively concentrated market structures ($n < 9$), the large gain in consumer surplus offsets the loss in profit resulting in a net welfare gain. However, for $n \geq 9$, the smaller gain in consumer surplus no longer offsets the loss in profits and there is a net loss in welfare. This is a rather counter-intuitive result: even though speculation is "pro-competitive" on average for large n in the sense that average price is lower, average welfare is lower than it would be in the absence of speculation. A box-and-whisker plot of the welfare effect is plotted in Fig. 11.9, where the skewness of the effects of speculation on welfare is evident: although the median change in welfare due to speculation is positive, there are relatively few periods with large welfare losses due to speculation. These welfare losses tend to occur in periods in which a stockout occurs (the C equilibrium). Given the state of demand, a_t, price is lower in these periods relative to the Cournot outcome. Although this generates benefits to consumers, much of this lower price is due to speculative sales, not due to an increase in output by producers. As producers see a lower price at the same time that they are lowering output, the loss in profits they see are larger than the gain that flows to consumers, resulting in a net welfare loss. In a sense, this welfare loss is due to storage costs as the “excess” loss of producer profit is compensating speculators for their storage costs incurred when they carry stocks. Even though speculators earn zero profit on average, their storage costs are essentially coming out of producer profit in periods in which stocks are sold.

11.6 Conclusion

By examining a dynamic game in which oligopolistic producers are faced with competitive speculators who can purchase, store and resell their output, we have seen that predictions of oligopoly theory can be substantially affected. In particular, the presence of speculators leads to more competitive behaviour by producers resulting in a reduction in the average price as compared to what occurs in the absence of speculators. In a computed example, we see that this effect can be quite large, on the order of 20 %. Conversely, speculators can have a substantial anti-competitive effect in the case of monopoly, where the attempts to deter speculative purchases leads to higher prices.

In contrast to studies of speculation in markets with competitive production, we find that the welfare effects of speculation are ambiguous in the oligopoly setting. Both consumer gains and producer profits are affected in a non-monotonic manner by speculative activity with the net effect either positive or negative. In our computed example, speculation improves mean social welfare when production is relatively concentrated, but reduces mean social welfare for less concentrated market structures.

References

Başar T, Olsder GJ (1995) Dynamic noncooperative game theory, 2nd edn. Academic, New York

Deaton A, Laroque G (1992) On the behavior of commodity prices. Rev Econ Stud 59:1–23

Deaton A, Laroque G (1996) Competitive storage and commodity price dynamics. J Polit Econ 104(5):896–923

European Commission (2011) Tackling the challenges in commodity markets and on raw materials. Communication 25

Frankel JA, Rose AK (2010) Determinants of agricultural and mineral commodity prices. In: Fry R, Jones C, Kent C (eds) Inflation in an era of relative price shocks. Reserve Bank of Australia, Sydney, pp 1–51

Gilbert CL (2010) Speculative influences on commodity futures prices 2006–2008. UNCTAD Discussion Paper No. 197

Jones E, Oliphant T, Peterson P, et al (2001–) SciPy: open source scientific tools for python. http://www.scipy.org/

Judd KL (1998) Numerical methods in economics. The MIT Press, Cambridge

Kilian L (2008) The economic effects of energy price shocks. J Econ Lit 46:871–909

Kilian L (2009) Not all price shocks are alike: disentangling demand and supply shocks in the crude oil market. Am Econ Rev 99:1053–1069

Kilian L, Murphy DP (2014) The role of inventories and speculative trading in the global market for crude oil. J Appl Econ 29(3):454–478

McLaren J (1999) Speculation on primary commodities: the effects of restricted entry. Rev Econ Stud 66:853–871

Mitraille S, Thille H (2009) Monopoly behaviour with speculative storage. J Econ Dyn Control 33:1451–1468

Mitraille S, Thille H (2014) Speculative storage in imperfectly competitive markets. Int J Ind Organ 35(C):44–59

Newbery DM (1984) Commodity price stabilization in imperfect or cartelized markets. Econometrica 52(3):563–578

Newbery DMG, Stiglitz JE (1981) The theory of commodity price stabilization: a study in the economics of risk. Oxford University Press, Oxford

Smith JL (2009) World oil: market or mayhem? J Econ Perspect 23(3):145–164

Thille H (2006) Inventories, market structure, and price volatility. J Econ Dyn Control 30(7):1081–1104

US Senate (2006) The role of market speculation in rising oil and gas prices: a need to put the cop back on the beat. Staff Report prepared by the Permanent Subcommittee on Investigations of the Committee on Homeland Security and Governmental Affairs

Williams JC, Wright BD (1991) Storage and commodity markets. Cambridge University Press, Cambridge

Chapter 12
Evolutionary Stability of Dimorphic Population States

Dharini Hingu, K.S. Mallikarjuna Rao, and A.J. Shaiju

Abstract We consider a dimorphic population state, P, which is a convex combination of two Dirac measures δ_x and δ_y, in evolutionary games with a continuous strategy space. We first establish necessary and sufficient conditions for this dimorphic population state, P, to be a rest point of the associated replicator dynamics. We provide sufficient conditions for the replicator dynamics trajectory to converge to P when it originates from the line $L = \{\eta\delta_x + (1-\eta)\delta_y : 0 < \eta < 1\}$. If the trajectory emanates from a point outside L, then we derive sufficient conditions for the trajectory to converge to L in the special case where each point in L is a rest point. We have, also, obtained condition for the trajectory to stay away from the line L in the limit. Furthermore, main results are illustrated using examples.

Keywords Evolutionary games with continuous strategies • Replicator dynamics • Lyapunov stability • Dimorphic population state

MSC Classification Codes: 37B25, 91A05, 91A22, 92D25

12.1 Introduction

Evolutionary games can be studied either with a discrete (pure) strategy space or with a continuous (pure) strategy space. Evolutionary games with a discrete strategy space have been widely studied in literature (Smith 1982; Weibull 1995; Hofbauer and Sigmund 1988; Cressmann 2003; Sandhlom 2010). There are many

D. Hingu (✉) • A.J. Shaiju
Department of Mathematics, Indian Institute of Technology Madras, Sardar Patel Road, Adayar, Chennai 600 036, Tamil Nadu, India
e-mail: dharinihingu@gmail.com; ajshaiju@iitm.ac.in

K.S. Mallikarjuna Rao
Industrial Engineering and Operations Research, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, Maharashtra, India
e-mail: mallik.rao@iitb.ac.in

© Springer International Publishing Switzerland 2016

F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_12

results available regarding the evolutionary stability of population states for these evolutionary games. Moreover, these results are studied with respect to several dynamics (Weibull 1995; Cressmann 2003; Sandhlom 2010). These results connect the static stability of population states with their dynamic stability.

The theory of evolutionary games with a continuous strategy space was initiated by Bomze and Pötscher (1989) and Bomze (1990), who introduced the concepts of uninvadable and strongly uninvadable strategies. This theory was further developed by Oechssler and Riedel (2001, 2002). In Oechssler and Riedel (2001), the authors proved that an uninvadable monomorphic population state is Lyapunov stable with respect to the replicator dynamics with the underlying topology as the variational (or strong) topology. In Oechssler and Riedel (2002), it is proved that, for a doubly symmetric game with compact strategy space and continuous payoff function, an evolutionarily robust population state is Lyapunov stable with respect to the replicator dynamics with the underlying topology as the weak topology.

The concept of a neighborhood attracting strategy was developed by Cressman (2005). A population state, Q, is called a neighborhood attracting strategy when a trajectory with the initial population state (whose support is close to the support of Q) converges to Q in the weak topology. One of the main results in Cressman (2005) is that a dimorphic neighborhood superior population state with finite support is neighborhood attracting. The book by Vincent and Brown (2005) is one of the latest references in the study of evolutionary games with continuous strategy space. In Shaiju and Bernhard (2009), it is proved that an evolutionarily robust monomorphic population state is weakly attracting.

In this paper, we first provide necessary and sufficient conditions for a dimorphic population state, $P = \alpha\, \delta_x + (1-\alpha)\, \delta_y$, to be a rest point of the replicator dynamics. Then we give sufficient conditions for this dimorphic population state to be stable when the initial population state is from the line L. We also provide with sufficient conditions for the trajectory to converge to the line L, when the initial population state is not from L. In this case we also give sufficient conditions for the trajectory to move away from the line L. The underlying topology here is the variational topology, i.e., the topology is defined using the variational norm (see the text before Definition 3).

The rest of the paper is arranged as follows. Section 12.2 covers the basic game getting along with the definitions of replicator dynamics and certain static and dynamic stability concepts. Section 12.3 deals with all the results in the previous paragraph. We end with some remarks in Sect. 12.4.

12.2 Preliminaries

Let S be a Borel set in $\mathbb{R}^n$. Consider the symmetric two-player game, G, with the pure strategy set S and the payoff function $u : S \times S \to \mathbb{R}$. We assume that the payoff function is bounded and Borel measurable. Here, S is equipped with the Borel σ-algebra $\mathscr{B}$. The measurability of $u(z, w)$ is with respect to this σ-algebra. By the symmetry of the game, we mean that Player I receives $u(z, w)$ when Player I plays $z \in S$ and Player II plays $w \in S$, whereas, Player II receives $u(w, z)$.

We follow the population interpretation of evolutionary games. A population state P is a probability measure on the measurable space $(S, \mathscr{B})$. Let Δ denote the set of all the population states. The average payoff of population P against population Q is given by,

$$E(P, Q) = \int_S \int_S u(z, w)\ Q(dw)\ P(dz).$$

We now introduce some (static) stability concepts such as, evolutionary stable strategy (ESS), uninvadable strategy and strongly uninvadable strategy (see Bomze and Pötscher (1989) for more details).

Definition 1. A population state P is called an *evolutionary stable strategy* (ESS) if for every "mutation" $Q \neq P$, there exists $\epsilon(Q) > 0$, such that, for all $0 < \eta \leq \epsilon(Q)$,

$$E(P, (1-\eta)P + \eta Q) > E(Q, (1-\eta)P + \eta Q).$$

The number $\epsilon(Q)$ is called invasion barrier corresponding to the mutation Q.

Definition 2. A population state P is called *uninvadable* if the invasion barrier $\epsilon(Q)$ can be chosen independent of Q.

Note that, for $R = (1-\eta)P + \eta Q$, from the definition of ESS, we have

$$E(P, R) = (1-\eta)E(P, R) + \eta E(P, R) > (1-\eta)E(P, R) + \eta E(Q, R) = E(R, R)$$

if $\eta < \epsilon(Q)$.

When the set of pure strategies is finite, then a neighborhood of P is completely characterized by the points R for η sufficiently small. However, this is not the case with infinite strategy space. One can define many topologies and in this article, we consider the topology induced by the variational (strong) norm i.e., the variational (or strong) topology. The variational norm of a probability measure P is given by

$$\|P\| = 2 \sup_{B \in \mathscr{B}} |P(B)|.$$

Thus the distance between two probability measures P and Q is given by

$$\|P - Q\| = 2 \sup_{B \in \mathscr{B}} |P(B) - Q(B)|.$$

Definition 3. A population state P is called *strongly uninvadable* if there is an $\epsilon > 0$ such that for all population states $R \neq P$ with $\|R - P\| \leq \epsilon$, we have

$$E(P, R) > E(R, R).$$

We now introduce replicator dynamics. For this, we consider a population evolving over time. Recall that the state of the population is given by a probability measure P on $(S, \mathscr{B})$.

The success (or lack of success) of a strategy $z \in S$, against a strategy, $w \in S$ is given by,

$$\sigma(z, w) = u(z, w) - u(w, w).$$

The average success (or lack of success) of a strategy $z \in S$, if the population state is Q, is given by,

$$\sigma(z, Q) := \int_S u(z, w)\ Q(dw) - \int_S \int_S u(z, w) Q(dw) Q(dz) = E(\delta_z, Q) - E(Q, Q).$$

The replicator dynamics is defined using the notion (similar to that in case of games with a discrete pure strategy space) that the relative increment in the frequency of strategies in a set $B \in \mathscr{B}$ is given by the average success of strategies in B. That is, for every $B \in \mathscr{B}$,

$$Q'(t)(B) = \int_B \sigma(z, Q(t))\ Q(t)(dz) \tag{12.1}$$

with the initial population state, $Q(0) = Q_0$.

The replicator dynamics may be written conveniently as : $Q'(t) = F(Q(t))$, where $F(Q(t))$ is the signed measure which is absolutely continuous w.r.t. $Q(t)$ with density $\sigma(\cdot, Q(t))$.

Theorem 1 (Oechssler and Riedel (2001)). *If the payoff function u is bounded, then the replicator dynamics is well-defined, in the sense that,* (12.1) *admits a unique solution.*

We now introduce some dynamic stability concepts, such as, Lyapunov stability, strongly attracting strategy and weakly attracting strategy.

Recall that, a rest point of replicator dynamics is a population state P such that $F(P) = 0$.

Definition 4. P is called *Lyapunov stable* if for all $\epsilon > 0$, there exists an $\eta > 0$ such that, for all $t > 0$

$$\|Q(0) - P\| < \eta \ \Rightarrow \ \|Q(t) - P\| < \epsilon.$$

Definition 5. P is called *strongly attracting* if $Q(t)$ converges to P strongly, whenever $Q(0)$ is close (strongly) to P.

Definition 6. P is called *weakly attracting* if $Q(t)$ converges to P weakly, whenever $Q(0)$ is close (weakly) to P.

Recall that a sequence of probability measures P_n on S converges weakly to a probability measure P on S if for every bounded continuous function $f : S \to \mathbb{R}$, we have

$$\int_S f(x)P_n(dx) \to \int_S f(x)P(dx)$$

as $n \to \infty$. We will not discuss the metric defining the weak convergence, instead we refer to Borkar (1995) for details.

In Oechssler and Riedel (2001), sufficient conditions are provided for a monomorphic population state $Q^* = \delta_x$ to be Lyapunov stable and weakly attracting. To be precise, the following result is proved in Oechssler and Riedel (2001).

Theorem 2 (Oechssler and Riedel (2001)). *If $Q^* = \delta_x$ is an uninvadable, monomorphic population state, then Q^* is Lyapunov stable. Moreover, if u is continuous then Q^* is weakly attracting.*

As mentioned in the introduction, the objective of this paper is to study conditions under which a dimorphic population state, $P = \alpha\delta_x + (1-\alpha)\delta_y$ $(x \neq y; 0 < \alpha < 1)$, is stable in the strong topology. This will be done in the next section.

12.3 Main Results

In order to investigate the evolutionary stability of the dimorphic population state,

$$P = \alpha\,\delta_x + (1-\alpha)\,\delta_y, \qquad x \neq y, \quad 0 < \alpha < 1 \tag{12.2}$$

we first derive necessary and sufficient conditions for it to be a rest point of the replicator dynamics.

Lemma 1. *For the dimorphic population state P given in (12.2), the following are equivalent:*

(i) *P is a rest point of the replicator dynamics.*
(ii) $\alpha\sigma(y,x) = (1-\alpha)\sigma(x,y)$.
(iii) *Either $\sigma(x,y) = \sigma(y,x) = 0$ or $\alpha = \frac{\sigma(x,y)}{\sigma(x,y)+\sigma(y,x)} \in (0,1)$.*

Remark 1. Note that the second and third parts of the lemma enforces that both $\sigma(x,y)$ and $\sigma(y,x)$ have same sign. That is, the success of strategy x against y as well as the success of strategy y against x have same sign. Later we show that, for a dimorphic state P to be limit of a replicator trajectory, the success of strategy x against y as well as the success of strategy y against x must be positive.

Proof. First we will prove that (i) $\iff$ (ii). Note that P is a rest point of the replicator dynamics if and only if

$$F(P)(B) = \int_B \sigma(z,P)P(dz) = 0$$

for every Borel set $B \subseteq S$. Since B is arbitrary, this is equivalent to

$$\sigma(\cdot, P) = 0 \text{ a.e. } P$$

Since P is supported on $\{x, y\}$, this is equivalent to the fact that $\sigma(x, P) = 0 = \sigma(y, P)$. This implies and is implied by

$$E(\delta_x, P) = E(\delta_y, P) = E(P, P).$$

A simple algebra reveals that this is equivalent to

$$\alpha\sigma(y,x) = (1-\alpha)\sigma(x,y),$$

completing the proof of (i) $\iff$(ii).

Note that (ii) can be written as

$$\alpha[\sigma(x,y) + \sigma(y,x)] = \sigma(x,y).$$

It is obvious from this that (ii) and (iii) are equivalent, completing the proof of the lemma. □

The Lyapunov stability of the population state P means that, if we start from a population state $Q(0)$ near P, then the resulting trajectory $Q(t)$ with respect to the replicator dynamics will also stay near P. We next try to understand when will the population state $Q(0)$ be close to P. To this end, we first prove the following lemma.

Lemma 2. *If* $Q(0)(\{x\}) = \beta > 0$ *and* $Q(0)(\{y\}) = 0$ *then,*

$$\|Q(0) - P\| \geq 2(1-\alpha).$$

Proof. Let $Q(0) = \beta\ \delta_x + (1-\beta)\ R$, where $R \in \Delta$ with $R(\{x\}) = R(\{y\}) = 0$. Then, by definition,

$$\begin{aligned}\|Q(0) - P\| &= 2 \sup_{B\in\mathscr{B}}\{|Q(0)(B) - P(B)|\}\\ &\geq 2\ |Q(0)(\{y\}) - P(\{y\})|\\ &= 2(1-\alpha).\end{aligned}$$

□

Similarly, we can prove the following lemma:

Lemma 3. *If* $Q(0)(\{y\}) = \gamma > 0$ *and* $Q(0)(\{x\}) = 0$ *then,*

$$\|Q(0) - P\| \geq 2\alpha.$$

In view of the above two lemmas, we note that any point in an arbitrarily small neighborhood of P must have positive weights on both x and y. In fact, we can say more.

Theorem 3. *Let P be as in (12.2) and $Q(0) = \beta\,\delta_x + \gamma\,\delta_y + (1-\beta-\gamma)\,R$, where $R \in \Delta$ with $R(\{x\}) = R(\{y\}) = 0$. Then,*

$$\|Q(0) - P\| \;\leq\; 2\;\max\{|\beta-\alpha|, |\gamma-(1-\alpha)|, 1-\beta-\gamma\}.$$

Proof. By the definition of variational norm, it follows that

$$\begin{aligned}
\|Q(0) - P\| &= 2\sup_{B\in\mathscr{B}} |Q(0)(B) - P(B)| \\
&\leq 2\max\Big\{|Q(0)(\{x\}) - P(\{x\})|,\ |Q(0)(\{y\}) - P(\{y\})|,\ |Q(0)(\{x,y\}) - P(\{x,y\})|, \\
&\qquad |Q(0)(S\smallsetminus\{x\}) - P(S\smallsetminus\{x\})|,\ |Q(0)(S\smallsetminus\{y\}) - P(S\smallsetminus\{y\})|, \\
&\qquad |Q(0)(S\smallsetminus\{x,y\}) - P(S\smallsetminus\{x,y\})|\Big\} \\
&= 2\max\{|\beta-\alpha|, |\gamma-(1-\alpha)|, |\beta+\gamma-1|, |\gamma+(1-\beta-\gamma)-(1-\alpha)|, \\
&\qquad |\beta+(1-\beta-\gamma)-\alpha|, |1-\beta-\gamma-0|\} \\
&= 2\max\{|\beta-\alpha|, |\gamma-(1-\alpha)|, 1-\beta-\gamma\}
\end{aligned}$$

which concludes the proof. □

Remark 2. From Theorem 3, it follows that the initial population state $Q(0)$ can be arbitrarily close to P only if it is of the form, $Q(0) = \beta\,\delta_x + \gamma\,\delta_y + (1-\beta-\gamma)\,R$ where, $R \in \Delta$ with $R(\{x\}) = R(\{y\}) = 0$ and $0 < \beta+\gamma \leq 1$. This leads us to analyze the stability of P in two different cases. The first case is when $\beta+\gamma = 1$; that is, the initial population state $Q(0)$ is from the line $L = \{Q \in \Delta : Q = \eta\,\delta_x + (1-\eta)\,\delta_y;\ \text{for some}\ 0 < \eta < 1\}$. The second case is $0 < \beta+\gamma < 1$.

12.3.1 Trajectory $Q(t)$ with $Q(0) \in L$

Since $Q(t)$ is absolutely continuous with respect to $Q(0)$ (follows from the replicator equation), we must have $Q(t) \in L$ if initially it lies on L. Thus, $Q(t)$ will be of the form

$$Q(t) = \beta(t)\delta_x + (1-\beta(t))\delta_y$$

with $\beta(0) = \beta$, when $Q(0) = \beta\delta_x + (1-\beta)\delta_y$ and $0 < \beta < 1$. Also note that $\beta(t) \in (0, 1)$ for every t. If $\beta = \alpha$, then $Q(t) \equiv Q(0)$. Thus to avoid the trivial case we assume, in the sequel, that $\beta \neq \alpha$.

Substituting the form of $Q(t)$ in the replicator equation, we obtain a differential equation for $\beta(t)$, which is given by

$$\beta'(t) = \beta(t)\, \sigma(x, Q(t)); \qquad \beta(0) = \beta \tag{12.3}$$

We will now compute the value of $\sigma(x, Q(t))$. It is easy to see that

$$E(\delta_x, Q(t)) = \int_S u(x, z)Q(t)(dz) = \beta(t)u(x, x) + (1 - \beta(t))u(x, y)$$

and

$$E(\delta_y, Q(t)) = \int_S u(y, z)Q(t)(dz) = \beta(t)u(y, x) + (1 - \beta(t))u(y, y).$$

Using the above two expressions, we get

$$\begin{aligned} E(Q(t), Q(t)) = \beta(t)\Big[\beta(t)u(x, x) + (1 - \beta(t))u(x, y)\Big] \\ + (1 - \beta(t))\Big[\beta(t)u(y, x) + (1 - \beta(t))u(y, y)\Big] \end{aligned}$$

From this, we obtain

$$\begin{aligned} \sigma(x, Q(t)) &= E(\delta_x, Q(t)) - E(Q(t), Q(t)) \\ &= (1 - \beta(t))\,\{\beta(t)u(x, x) + (1 - \beta(t))u(x, y) - \beta(t)u(y, x) - (1 - \beta(t))u(y, y)\} \\ &= (1 - \beta(t))\,\{(1 - \beta(t))\sigma(x, y) - \beta(t)\sigma(y, x)\} \end{aligned}$$

From Lemma 1, we know that P is a rest point if either $\sigma(x, y) = \sigma(y, x) = 0$ or

$$\alpha = \frac{\sigma(x, y)}{\sigma(x, y) + \sigma(y, x)} \in (0, 1).$$

In the first case, we will have $\sigma(x, Q(t)) = 0$ and hence, from (12.3), we have $\beta(t) \equiv \beta$.

In the second case, we have

$$\sigma(x, Q(t)) = (1 - \beta(t))(\alpha - \beta(t))\frac{\sigma(x, y)}{\alpha}.$$

Substituting this in (12.3), we obtain

$$\beta'(t) = \beta(t)(1-\beta(t))(\alpha-\beta(t))\frac{\sigma(x,y)}{\alpha}.$$

which can be rewritten as

$$\frac{\beta'(t)}{\beta(t)(1-\beta(t))(\alpha-\beta(t))} = \frac{\sigma(x,y)}{\alpha} \tag{12.4}$$

Using partial fractions, we write the L.H.S. as

$$\frac{\beta'(t)}{\beta(t)(1-\beta(t))(\alpha-\beta(t))} = \frac{\beta'(t)}{\alpha\beta(t)} - \frac{\beta'(t)}{(1-\alpha)(1-\beta(t))} + \frac{\beta'(t)}{\alpha(1-\alpha)(\alpha-\beta(t))}.$$

Integrating (12.4) from 0 to t, we get, after some simplifications,

$$(1-\alpha)\ln\left(\frac{\beta(t)}{\beta}\right) + \alpha\ln\left(\frac{1-\beta(t)}{1-\beta}\right) + \ln\left(\frac{\alpha-\beta}{\alpha-\beta(t)}\right) = (1-\alpha)\sigma(x,y)t.$$

Simplifying this further, we obtain

$$\frac{(\beta(t))^{1-\alpha}(1-\beta(t))^{\alpha}}{(\alpha-\beta(t))} = \frac{\beta^{1-\alpha}(1-\beta)^{\alpha}}{(\alpha-\beta)}\exp\Big((1-\alpha)\sigma(x,y)t\Big) \tag{12.5}$$

It can be observed from (12.5) that, the stability of P will depend on the sign of $\sigma(x,y)$ as stated in the following theorem.

Theorem 4. *Let $P = \alpha\delta_x + (1-\alpha)\delta_y$ be a rest point of the replicator dynamics and the initial population state be of the form $Q(0) = \beta\delta_x + (1-\beta)\delta_y$; $(0 < \beta < 1)$. Then,*

1. *the trajectory $Q(t) = Q(0)$ for all t when $\sigma(x,y) = 0$,*
2. *the trajectory $Q(t)$ converges to P when $\sigma(x,y) > 0$, and*
3. *the trajectory $Q(t)$ either converges to δ_x or δ_y when $\sigma(x,y) < 0$.*

Proof. **Case 1.** Suppose $\sigma(x,y) = 0$.
From Lemma 1, we have $\sigma(y,x) = 0$ and hence any point on line L is a rest point of replicator dynamics. So $Q(t) = Q(0)$ for every t.
Case 2. Suppose $\sigma(x,y) > 0$.
It is clear that the R.H.S. of (12.5) tends to ∞ as $t \to \infty$. Thus

$$\frac{(\beta(t))^{1-\alpha}(1-\beta(t))^{\alpha}}{(\alpha-\beta(t))} \to \infty.$$

Since $\beta(t)$ is bounded, the numerator is always bounded. Thus the only way, the above can happen, is that $\alpha - \beta(t) \to 0$. Hence $Q(t)$ converges to P as $t \to \infty$.

Case 3. Suppose $\sigma(x, y) < 0$.
From (12.5), we obtain

$$\frac{(\beta(t))^{1-\alpha}(1-\beta(t))^{\alpha}}{(\alpha-\beta(t))} \to 0.$$

Since $\beta(t)$ is bounded, the denominator can not go to ∞. Thus, for this to happen, we must have

$$(\beta(t))^{1-\alpha}(1-\beta(t))^{\alpha} \to 0.$$

Hence, $\beta(t) \to 1$ or 0. Consequently the trajectory $Q(t)$ converges to either δ_x or δ_y. □

The following example illustrates the above result.

Example 1 (Shaiju and Bernhard (2009)). Let $S = [0, 1]$ and $\lambda \in (0, \infty)$. Let the payoff function be defined as,

$$u(z, w) = \max\{z - w, \lambda(w - z)\} \qquad \forall\, z,\ w \in S.$$

Let δ_0 and δ_1 denote the Dirac measures concentrated on the points $x = 0$ and $y = 1$ respectively. In this case,

$$\sigma(x, y) = \sigma(0, 1) = u(0, 1) - u(1, 1) = \lambda$$

and similarly, $\sigma(y, x) = \sigma(1, 0) = 1$. Thus, by Lemma 1, $P = \alpha\delta_0 + (1-\alpha)\delta_1$ is a rest point of the replicator dynamics with

$$\alpha = \frac{\lambda}{1+\lambda}.$$

Now, consider the replicator dynamics with the initial population state $Q(0) \in L$. Then, as a consequence of Theorem 4, the trajectory $Q(t)$ converges to P since $\sigma(x, y) = \lambda > 0$.

12.3.2 Trajectory $Q(t)$ with $Q(0) \notin L$

In this case the initial population state is of the form

$$Q(0) = \beta\delta_x + \gamma\delta_y + (1-\beta-\gamma)R,$$

where $0 < \beta + \gamma < 1$ and $R \in \Delta$ with $R(\{x\}) = R(\{y\}) = 0$. The trajectory $Q(t)$, being absolutely continuous w.r.t. $Q(0)$, is of the form

$$Q(t) = \beta(t)\delta_x + \gamma(t)\delta_y + (1 - \beta(t) - \gamma(t))R(t),$$

where $\beta(t)$ and $\gamma(t)$ are solutions of the following system of differential equations (obtained by substituting $B = \{x\}$ and $B = \{y\}$ respectively in the replicator dynamics (12.1)),

$$\beta'(t) = \beta(t)\,\sigma(x, Q(t)) \quad \beta(0) = \beta \tag{12.6}$$

$$\gamma'(t) = \gamma(t)\,\sigma(y, Q(t)) \quad \gamma(0) = \gamma. \tag{12.7}$$

Further note that, $R(0) = R$ and $R(t)$ is absolutely continuous w.r.t. $R(0)$. As in the previous subsection, $\beta(t) \in (0, 1)$ and $\gamma(t) \in (0, 1)$ for all t.

We first analyze the case $\sigma(x, y) = \sigma(y, x) = 0$ where, each point on the line L will be a rest point of the replicator dynamics. To this end, we now compute $E(\delta_x, Q(t))$, $E(\delta_y, Q(t))$ and $E(Q(t), Q(t))$.

$$\begin{aligned} E(\delta_x, Q(t)) &= \int_S u(x, z)\; Q(t)(dz) \\ &= \beta(t)u(x, x) + \gamma(t)u(x, y) + (1 - \beta(t) - \gamma(t))E(\delta_x, R(t)). \end{aligned}$$

Since $\sigma(x, y) = 0$, we have $u(x, y) = u(y, y)$ and hence

$$E(\delta_x, Q(t)) = \beta(t)u(x, x) + \gamma(t)u(y, y) + (1 - \beta(t) - \gamma(t))\; E(\delta_x, R(t)).$$

Similarly, by using the fact that $\sigma(y, x) = 0$, we have

$$E(\delta_y, Q(t)) = \beta(t)u(x, x) + \gamma(t)u(y, y) + (1 - \beta(t) - \gamma(t))\; E(\delta_y, R(t)).$$

Using the above expressions for $E(\delta_x, Q(t))$ and $E(\delta_y, Q(t))$ we get,

$$\begin{aligned} E(Q(t), Q(t)) = \beta(t)&\Big[\beta(t)u(x, x) + \gamma(t)u(y, y) + (1 - \beta(t) - \gamma(t))E(\delta_x, R(t))\Big] \\ &+ \gamma(t)\Big[\beta(t)u(x, x) + \gamma(t)u(y, y) + (1 - \beta(t) - \gamma(t))E(\delta_y, R(t))\Big] \\ &+ (1 - \beta(t) - \gamma(t))\Big[\beta(t)E(R(t), \delta_x) + \gamma(t)E(R(t), \delta_y) \\ &\qquad + (1 - \beta(t) - \gamma(t))E(R(t), R(t))\Big] \end{aligned}$$

By using the expressions for $E(\delta_x, Q(t))$ and $E(Q(t), Q(t))$ we get,

$$\begin{aligned}
\sigma(x, Q(t)) &= E(\delta_x, Q(t)) - E(Q(t), Q(t)) \\
&= \beta(t)u(x,x) + \gamma(t)u(y,y) + (1-\beta(t)-\gamma(t))\, E(\delta_x, R(t)) \\
&\quad - \beta(t)\Big[\beta(t)u(x,x) + \gamma(t)u(y,y) + (1-\beta(t)-\gamma(t))\, E(\delta_x, R(t))\Big] \\
&\quad - \gamma(t)\Big[\beta(t)u(x,x) + \gamma(t)u(y,y) + (1-\beta(t)-\gamma(t))\, E(\delta_y, R(t))\Big] \\
&\quad - (1-\beta(t)-\gamma(t))\Big[\beta(t)\, E(R(t), \delta_x) + \gamma(t)\, E(R(t), \delta_y) \\
&\quad + (1-\beta(t)-\gamma(t))\, E(R(t), R(t))\Big].
\end{aligned}$$

After rearranging the terms it becomes,

$$\begin{aligned}
\sigma(x, Q(t)) = (1-\beta(t)-\gamma(t))\Big[&-\beta(t)\sigma(R(t), x) - \gamma(t)\sigma(R(t), y) \\
&+ (1-\beta(t))\sigma(x, R(t)) - \gamma(t)\sigma(y, R(t))\Big]
\end{aligned} \tag{12.8}$$

where

$$\begin{aligned}
\sigma(R(t), x) &= E(R(t), \delta_x) - u(x,x), \\
\sigma(R(t), y) &= E(R(t), \delta_y) - u(y,y), \\
\sigma(x, R(t)) &= E(\delta_x, R(t)) - E(R(t), R(t)), \text{ and} \\
\sigma(y, R(t)) &= E(\delta_y, R(t)) - E(R(t), R(t)).
\end{aligned}$$

In a similar fashion, we can obtain

$$\begin{aligned}
\sigma(y, Q(t)) = (1-\beta(t)-\gamma(t))\Big[&-\beta(t)\sigma(R(t), x) - \gamma(t)\sigma(R(t), y) \\
&- \beta(t)\sigma(x, R(t) + (1-\gamma(t))\sigma(y, R(t))\Big]
\end{aligned} \tag{12.9}$$

Let

$$c(t) = \min\Big\{-\sigma(R(t), x),\ -\sigma(R(t), y),\ \sigma(x, R(t)),\ \sigma(y, R(t))\Big\} \tag{12.10}$$

$$C(t) = \max\Big\{-\sigma(R(t), x),\ -\sigma(R(t), y),\ \sigma(x, R(t)),\ \sigma(y, R(t))\Big\}. \tag{12.11}$$

The following theorem describes some conditions for which the trajectory $Q(t)$ converges to a point on the line L.

Theorem 5. *Let the dimorphic population state P, as given in* (12.2)*, be a rest point of the replicator dynamics with* $\sigma(x,y) = \sigma(y,x) = 0$ *and the initial population state* $Q(0)$ *be of the form*

$$Q(0) = \beta\delta_x + \gamma\delta_y + (1-\beta-\gamma)R \ ; \quad 0 < \beta+\gamma < 1$$

where $R \in \Delta$ *with* $R(\{x\}) = R(\{y\}) = 0$*. If*

$$\int_0^\infty [2c(t) - C(t)]\, dt = \infty, \tag{12.12}$$

where $c(t)$ *and* $C(t)$ *are as defined by* (12.10) *and* (12.11) *respectively, then the trajectory* $Q(t)$ *converges to the line L.*

Proof. In order to prove the theorem, we consider lower bounds of $\sigma(x, Q(t))$ and $\sigma(y, Q(t))$. Using $c(t)$ and $C(t)$, we have

$$\begin{aligned}
\sigma(x,Q(t)) &\geq (1-\beta(t)-\gamma(t))\Big[\beta(t)c(t) + \gamma(t)c(t) + (1-\beta(t))c(t) - \gamma(t)C(t)\Big] \\
&= (1-\beta(t)-\gamma(t))\Big[c(t) - \gamma(t)\{C(t)-c(t)\}\Big] \\
&\geq (1-\beta(t)-\gamma(t))\Big[c(t) - \{C(t)-c(t)\}\Big] \\
&= (1-\beta(t)-\gamma(t))\Big[2c(t) - C(t)\Big].
\end{aligned}$$

Here, we have also used the fact that $\gamma(t) < 1$. Similarly,

$$\begin{aligned}
\sigma(y,Q(t)) &\geq (1-\beta(t)-\gamma(t))\Big[\beta(t)c(t) + \gamma(t)c(t) - \beta(t)C(t) + (1-\gamma(t))c(t)\Big] \\
&= (1-\beta(t)-\gamma(t))\Big[c(t) - \beta(t)\{C(t)-c(t)\}\Big] \\
&\geq (1-\beta(t)-\gamma(t))\Big[c(t) - \{C(t)-c(t)\}\Big] \\
&= (1-\beta(t)-\gamma(t))\Big[2c(t) - C(t)\Big].
\end{aligned}$$

Now let $p(t) = \beta(t) + \gamma(t) \in (0,1)$. Using these two lower bounds and Eqs. (12.6) and (12.7), we have the following differential inequality

$$p'(t) \geq p(t)(1-p(t))\big[2c(t) - C(t)\big]$$

with the initial condition $p_0 = \beta + \gamma$. This equation can be rewritten as

$$\frac{p'(t)}{p(t)} + \frac{p'(t)}{1-p(t)} \geq 2c(t) - C(t)$$

A little bit of analysis will yield the estimate

$$\frac{p(t)}{1-p(t)} \geq \exp\left\{\int_0^t (2c(s) - C(s))ds\right\}$$

Now the assumption (12.12) implies that $p(t) = \beta(t) + \gamma(t) \to 1$ as $t \to \infty$. Thus, the trajectory $Q(t)$ converges to the line L. □

Next we look at an example which illustrates the above result.

Example 2. Let $S = [0, 1]$ and the payoff function be defined as

$$u(z, w) = \begin{cases} (1-z)(1-w) : & z,\ w < \frac{1}{2} \\ w(1-z) : & z \leq \frac{1}{2},\ \ w \geq \frac{1}{2} \\ z(1-w) : & z \geq \frac{1}{2},\ \ w \leq \frac{1}{2} \\ zw : & z,\ w > \frac{1}{2} \end{cases}$$

Let $x = 0$ and $y = 1$. Then,

$$\sigma(x, y) = \sigma(0, 1) = u(0, 1) - u(1, 1) = 0 \text{ and}$$

$$\sigma(y, x) = \sigma(1, 0) = u(1, 0) - u(0, 0) = 0.$$

Thus, each point on the line L is a rest point of the replicator dynamics.

Consider, the initial population state of the form

$$Q(0) = \beta\delta_0 + \gamma\delta_1 + (1 - \beta - \gamma)\delta_{1/3}; \quad 0 < \beta + \gamma < 1.$$

Then the trajectory $Q(t)$ will be of the form

$$Q(t) = \beta(t)\delta_0 + \gamma(t)\delta_1 + (1 - \beta(t) - \gamma(t))\delta_{1/3}.$$

That is, $R(t) \equiv \delta_{1/3}$.

Now for this $R(t)$ we first evaluate the following to obtain $c(t)$ and $C(t)$.

$$-\sigma(R(t), x) = -\sigma\left(\frac{1}{3}, 0\right) = u(0, 0) - u\left(\frac{1}{3}, 0\right) = \frac{1}{3}$$

$$-\sigma(R(t), y) = -\sigma\left(\frac{1}{3}, 1\right) = u(1, 1) - u\left(\frac{1}{3}, 1\right) = \frac{1}{3}$$

$$\sigma(x, R(t)) = \sigma\left(0, \frac{1}{3}\right) = u\left(0, \frac{1}{3}\right) - u\left(\frac{1}{3}, \frac{1}{3}\right) = \frac{2}{9}$$

$$\sigma(y, R(t)) = \sigma\left(1, \frac{1}{3}\right) = u\left(1, \frac{1}{3}\right) - u\left(\frac{1}{3}, \frac{1}{3}\right) = \frac{2}{9}$$

This implies that

$$c(t) = \min\left\{\frac{1}{3}, \frac{2}{9}\right\} = \frac{2}{9} \text{ and}$$

$$C(t) = \max\left\{\frac{1}{3}, \frac{2}{9}\right\} = \frac{1}{3}$$

We observe that,

$$\int_0^\infty [2c(t) - C(t)]dt = \int_0^\infty \left[2\left(\frac{2}{9}\right) - \frac{1}{3}\right] dt = \int_0^\infty \frac{1}{9} dt = \infty.$$

Hence by Theorem 5, the trajectory $Q(t)$ converges to the line L.

We, now, consider the case where the trajectory will go away from the line. This, we accomplish in the following theorem.

Theorem 6. *Let the dimorphic population state P, as given in (12.2), be a rest point of the replicator dynamics with $\sigma(x,y) = \sigma(y,x) = 0$ and the initial population state $Q(0)$ be of the form*

$$Q(0) = \beta\delta_x + \gamma\delta_y + (1-\beta-\gamma)R \;\; ; \quad 0 < \beta + \gamma < 1$$

where $R \in \Delta$ with $R(\{x\}) = R(\{y\}) = 0$. If

$$\int_0^\infty [2C(t) - c(t)]\, dt = -\infty, \tag{12.13}$$

where $c(t)$ and $C(t)$ are as defined by (12.10) and (12.11) respectively, then the trajectory $Q(t)$ converges (or diverges) away from the line L.

Proof. Idea of the proof is exactly same as the previous theorem. Here, we consider upper bounds rather than the lower bounds.

Note that,

$$\begin{aligned}
\sigma(x, Q(t)) &\leq (1-\beta(t)-\gamma(t))\Big[\beta(t)C(t) + \gamma(t)C(t) + (1-\beta(t))C(t) - \gamma(t)c(t)\Big] \\
&= (1-\beta(t)-\gamma(t))\Big[C(t) + \gamma(t)\{C(t) - c(t)\}\Big] \\
&\leq (1-\beta(t)-\gamma(t))\Big[C(t) + C(t) - c(t)\Big] \\
&= (1-\beta(t)-\gamma(t))\Big[2C(t) - c(t)\Big].
\end{aligned}$$

Similarly,

$$\sigma(y, Q(t)) \le (1 - \beta(t) - \gamma(t))\Big[2C(t) - c(t)\Big].$$

Considering $p(t) = \beta(t) + \gamma(t) \in (0, 1)$, we have the following differential inequality

$$p'(t) \le p(t)(1 - p(t))\big[2C(t) - c(t)\big]$$

with the initial condition $p_0 = \beta + \gamma$. This equation can be rewritten as

$$\frac{p'(t)}{p(t)} + \frac{p'(t)}{1 - p(t)} \le 2C(t) - c(t)$$

A little bit of analysis will yield the estimate

$$\frac{p(t)}{1 - p(t)} \le \exp\left\{\int_0^t (2C(s) - c(s))ds\right\}$$

Now the assumption (12.13) implies that $p(t) = \beta(t) + \gamma(t) \to 0$ as $t \to \infty$. Thus, the trajectory $Q(t)$ goes away from the line L as $t \to \infty$. □

The above theorem can be illustrated using the next example.

Example 3. Let $S = [0, 1]$ and the payoff function be defined as:

$$u(z, w) = z(1 - z) + w(1 - w).$$

By Lemma 1, each point on the line joining δ_0 and δ_1 is a rest point of the associated replicator dynamics as, $\sigma(0, 1) = \sigma(1, 0) = 0$. We take $R = \delta_{1/2}$ and compute

$$\sigma(0, R(t)) = \sigma\left(0, \frac{1}{2}\right) = -\frac{1}{4},$$

$\sigma(1, R(t)) = -\frac{1}{4}$, $\sigma(R(t), 0) = \frac{1}{4}$ and $\sigma(R(t), 1) = \frac{1}{4}$.

This implies that, $C(t) = c(t) = -\frac{1}{4}$ and hence condition (12.13) is satisfied. Therefore, by Theorem 6, $\beta(t) + \gamma(t) \to 0$.

12.4 Concluding Remarks

In Sect. 12.3.2, we have discussed the convergence of the replicator dynamics trajectory when $Q(0) \notin L$. If $\sigma(x, y) = \sigma(y, x) = 0$, a set of sufficient conditions are given (Theorem 5) for the convergence of the trajectory to L and also to be away from the line L (Theorem 6). It would be nice if one can weaken these assumptions. Furthermore, the convergence of trajectories when $Q(0) \notin L$ and $\sigma(x,y) \neq 0$ is also an interesting question for further investigation. To this end, it is interesting to note that when we have a game G, in which $R(t) \equiv R(0)$ for all t, the system of differential equations for $\beta(\cdot)$ and $\gamma(\cdot)$ is equivalent to the replicator dynamics for a 3×3 game with the payoff matrix given by

$$U(R(0)) = \begin{pmatrix} u(x,x) & u(x,y) & E(\delta_x, R(0)) \\ u(y,x) & u(y,y) & E(\delta_y, R(0)) \\ E(R(0), \delta_x) & E(R(0), \delta_y) & E(R(0), R(0)) \end{pmatrix}.$$

In this case, the dimorphic population state $P = \alpha\delta_x + (1-\alpha)\delta_y$ will be stable when $(\alpha,\ 1-\alpha,\ 0)$ is a NSS (neutrally stable strategy) and it will be asymptotically stable when it is an ESS. If $R(t)$ is different from $R(0)$ then, a generalization of such ideas can be helpful to prove results regarding the stability of the dimorphic population states.

Acknowledgements The authors thank an anonymous referee for pointing out an error in an example in the preliminary version of this paper. We also acknowledge the support of NBHM through the project "Evolutionary Stability in Games with Continuous Action Spaces".

References

Bomze IM (1990) Dynamical aspects of evolutionary stability. Monatsh Math 110:189–206

Bomze IM, Pötscher BM (1989) Game theoretical foundations of evolutionary stability. Springer, Berlin/Heidelberg

Borkar VS (1995) Probability theory: an advanced course. Springer, New York

Cressmann R (2003) Evolutionary dynamics and extensive form games. MIT Press, London

Cressman R (2005) Stability of the replicator equation with continuous strategy space. Math Soc Sci 50:127–147

Hofbauer J, Sigmund K (1988) The theory of evolution and dynamical systems. Cambridge University Press, Cambridge. Paul Parey (1984) Evolutionstheorie und dynamische Systeme: Mathematische Aspekte der Selektion. Verlag, Hamburg

Maynard Smith J (1982) Evolution and theory of games. Cambridge University Press, Cambridge

Oechssler J, Riedel F (2001) Evolutionary dynamics on infinite strategy spaces. Econ Theory 17:141–162

Oechssler J, Riedel F (2002) On the dynamic foundation of evolutionary stability in continuous models. J Econ Theory 107:223–252

Sandhlom W (2010) Population games and evolutionary dynamics. MIT Press, Cambridge

Shaiju AJ, Bernhard P (2009) Evolutionarily robust strategies: two nontrivial examples and a theorem. In: Odile P, Vladimir G, Pierre B (eds) Advances in dynamic games and their applications. Annals of the International Society of Dynamic Games, vol 10. Birkhäuser, Boston, pp 1–19

Vincent T, Brown J (2005) Evolutionary game theory, natural selection and Darwanian dynamics. Cambridge University Press, Cambridge

Weibull J (1995) Evolutionary game theory. MIT Press, Cambridge

Chapter 13
A Game Theoretical Approach to Microbial Coexistence

Monica Abrudan, Li You, Kateřina Staňková, and Frank Thuijsman

Abstract Huge numbers of microbes coexist in almost all habitats of our planet. Their interactions are governed by complex mechanisms, where both competition for resources and toxin production play important roles. Our goal is to understand key mechanisms that lead to coexistence. In this chapter we study many possible scenarios of microbial interactions and we analyze whether or not they can lead to coexistence of species. To achieve this we implemented agent-based models that mimic local dynamics of microbes; initially well mixed microbes from different species interact in a grid with a regular structure. Among others, we show that the coexistence rate is negatively correlated with the number of neighbors of each cell in the grid. Another observation is that the order of selection of focal cells in the grid influences the coexistence rate.

Keywords Local dynamics • Bacterial strains • Agent based model • Game theory

MSC Classification Codes: 92D40, 91A22, 37B15

M. Abrudan
Faculty of Life Sciences, University of Manchester, Michael Smith Building, Dover Street, Manchester M13 9PL, UK

Institute of Biology, Sylvius Building, Sylviusweg 72, 2333 BE, Leiden University, Leiden, The Netherlands
e-mail: ma14@sanger.ac.uk

L. You • F. Thuijsman
Department of Data Science and Knowledge Engineering, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
e-mail: li.you@maastrichtuniversity.nl; f.thuijsman@maastrichtuniversity.nl

K. Staňková (✉)
Department of Data Science and Knowledge Engineering, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
e-mail: k.stankova@maastrichtuniversity.nl

© Springer International Publishing Switzerland 2016

F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_13

13.1 Introduction and Literature Review

Bacteria outnumber by far any other living organisms on Earth (Dykhuizen 1998). There are more bacteria on Earth than stars in the universe and it is thought that 90 % of the cells associated with a human body are not ours, but belong to microorganisms. With the help of new sequencing techniques, pyrosequencing and next generation sequencing, even environments considered sterile until a short time ago, like the uterus (Funkhouser and Bordenstein 2013) and mother milk (Hunt et al. 2011), were found to host microbes. Our faces are not clean either: recent research found that *Demodex* mites might be colonizing all of us (Thoemmes et al. 2014).

What is fascinating about bacteria is not just their astronomically high numbers, but also their very high species diversity. Two hundred and forty two healthy individuals were found to host more that 1200 different microbes (Consortium 2012) and a gram of soil can contain up to 8000 different operational taxonomic units (OTUs) (Delmont et al. 2011).

These amazing numbers raise an obvious question: How is the diversity maintained? Understanding the mechanisms that maintain microbial diversity could prove to be very beneficial to humans, as explained in the following examples: Firstly, equilibrium of commensal species could prevent invasion of pathogens, keeping ecosystems and humans healthy. Secondly, it was proved that the distribution of commensal species affects the homeostasis of the host. For example, in the mouse gut an abnormal ratio between *Firmicutes* and *Bacteroides* was found to be correlated to obesity (Ridaura et al. 2013).

Species interactions are thought to be a good proxy for understanding the mechanisms that drive diversity. This fact has lead ecologists to study diverse communities from the perspective of competition networks, mutualities networks, and food webs (Verhoef and Morin 2010). The topologies of the interaction matrices/networks are a key indicator of the species' dynamics in a given environment (Dunne et al. 2002; Bascompte et al. 2006; Montoya et al. 2006).

These observations in natural environments have led us to run a number of simulation experiments as a first attempt in trying to understand the underlying processes.

Bacterial interaction matrices are built using diverse experimental parameters, like growth rates, ability to produce toxins in pairwise interactions or "metabolic burdens" associated to expressing different traits

A classical experiment with two strains of *E. coli* (Chao and Levin 1981) showed that competition is regulated by the ability of the strains to produce toxins, as well as by their initial densities. Kerr et al. (2002) showed that equilibrium could be maintained between three strains of *E. coli*, which followed the setups of the Rock-Paper-Scissors game by careful control of their growth rates and of their sensitivity to toxins.

Taking this rationale further, the interactions inside a community of different strains of bacteria can be described through an $n \times n$ matrix $A = (a_{ij})$ (where n is the number of species), which describes the pairwise interactions between all species present. If entry a_{ij} is 1, then type i can kill type j, if it is 0, it cannot. We will refer to such a matrix as an interaction matrix.

While one can argue that in reality the entries of the interaction matrix should be expressed by the probabilities $p_{ij} \in [0, 1]$ at which strain i can kill strain j, it would be very difficult to derive these probabilities in laboratory measurements and therefore experimentalists confine themselves to discrete $0/1$ values. In this chapter, we will confine ourselves to this simplification as well, bearing in mind that the situation considered is a special case of the general case.

The structures of the interaction matrices of coexisting bacteria are very diverse

We will refer to the density of an interaction matrix as the proportion of 1's in that matrix.

Several studies focused on the ability of bacteria to inhibit other bacteria, that is, to secrete killing toxins. Vetsigian et al. (2011) characterized a set of 59 strains of *Streptomyces* and found that 43 % of the interactions assessed were cases of inhibition. In (Kinkel et al. 2013) it is reported that in seven communities of *Streptomyces* collected from seven different geographic locations, frequencies of inhibitions between sympatric isolates varied between 10 % and 33 %. In two different communities of *Streptomyces* collected from Leidse Hout in The Netherlands we found rates of inhibition of 33 % and 43 %, respectively. In communities of *E. coli* inhibition is less common, with only 30 % of *E. coli* strains producing colicin in natural environment (Riley and Gordon 1999). Seventy percent of the *E. coli* strains assessed were resistant to at least one colicin, while 30 % were resistant to all colicin produced. In a community of 25 strains of *Streptococcus pneumoniae* 22 % of the interactions between strains were cases of inhibition, where all strains were capable of producing toxins and all strains were vulnerable to at least one toxin. At a first glance, what is interesting about these interaction matrices is that they have very different densities. This data is summarized in Table 13.1. In all studies presented, the interaction matrices were constructed based on pairwise antagonistic interactions: in every community, each strain was tested whether it can secrete toxins that kill the other strains.

Table 13.1 Densities of different microbial interaction matrices

Species	Number of strains	Density of the interaction matrix
Streptomyces	59	0.43
Seven communities of *Streptomyces*	69	0.1–0.33
Streptomyces (unpublished data)	23	0.33
Streptomyces (unpublished data)	20	0.42
Streptomyces (unpublished data)	13	0.25
E. coli		0.07–0.21
E. coli ECOR collection	72	0.022
Streptococcus pneumoniae (unpublished data)	25	0.22

Results were determined experimentally

Quorum sensing often regulates microbial interactions

It is important to notice that most bacterial strains do not live in isolation. Bacteria can interact/communicate with one another using chemical signal molecules. This form of signalling is termed *quorum sensing* and enables bacteria to coordinate their behavior (Miller and Bassler 2001; Waters and Bassler 2005). As the environment for a bacteria population always changes rapidly, bacteria need to respond quickly in order to survive. These responses include adaptation to availability of nutrients, defence against other microorganisms which may compete for the same nutrients and the avoidance of toxic compounds potentially dangerous for the bacteria. As a result bacteria are able to monitor the environment for other bacteria and to alter behavior in response to changes in the numbers of strains present in a community (Williams et al. 2007). In this study we implemented a quorum sensing model (see Sect. 13.3.3).

The remainder of this chapter is composed as follows. In Sect. 13.2 our models and their implementation are explained. In Sect. 13.3 we discuss four case studies using different simulation setups. We end the chapter with remarks, conclusions and ideas for future work (Sect. 13.4).

13.2 Model

13.2.1 Basics

Let us consider a thought experiment: At the beginning of the experiment there are n bacterial strains and an equal number of cells of each strain, well mixed on a plate. We will denote the set of all strains by $N \stackrel{\text{def}}{=} \{1, \ldots, n\}$. We will represent the cells on the plate as a field, in two possible ways:

Fig. 13.1 An illustration of a square field, in which each cell has eight neighbors (*left*) and a field composed of hexagonal cells, in which each cell has six neighbors (*right*). In order to have the boundary conditions satisfied we assume that each type of field forms the surface of a torus. No cell is empty and cells with the same color belong to the same strain. In each figure only one focal cell (the *central one*) and its neighbors are depicted

- a squared field in which each cell has eight neighbors (Fig. 13.1, left);
- a field composed of hexagonal cells in which each cell has six neighbors (right) (Fig. 13.1, right).

We will assume that no cells are empty. Two different variants of the interactions will be considered. The initial model in which only simple interactions among cells will be taken into account, is introduced in Sect. 13.2.2, while a more advanced model, in which quorum sensing takes place, is introduced in Sect. 13.2.3.

13.2.2 Basic Model

Let us assume that at each time step of the experiment, one cell (the so-called **focal cell**) is randomly selected and interacts with another cell from its neighborhood (where the neighborhood contains six or eight cells). If the cell is of the type that can kill the neighbor, it will replace it, otherwise not. Killing the neighbor and replacing it (i.e., taking over that neighbor cell) represent one way of reproduction. Focal individuals that do not kill are assumed to reproduce in their own cell. No individual can kill itself or an individual of its own type, each individual $i \in N$ can either kill the strain $j \in N, j \neq i$ or not. We can represent the ability of the strains to kill each other by an **interaction matrix** of the following structure:

$$\begin{array}{c} \\ \text{type } 1 \\ \text{type } 2 \\ \vdots \\ \text{type } n \end{array} \begin{array}{c} \begin{array}{cccc} \text{type } 1 & \text{type } 2 & \dots & \text{type } n \end{array} \\ \begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ a_{21} & 0 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n\,1} & \dots & a_{n\,n-1} & 0 \end{pmatrix} \end{array}, \tag{13.1}$$

with

$$a_{ij} = \begin{cases} 1, & \text{if type } i \text{ can kill type } j \\ 0, & \text{otherwise,} \end{cases}$$

where $i \neq j$. The interactions given by matrix (13.1) can be equivalently illustrated in a directed graph with n vertices, in which a directed link from vertex i to vertex j corresponds to $a_{ij} = 1$ and there is no directed link from i to j when $a_{ij} = 0$. Note that $a_{ij} = 1$ does not imply that $a_{ji} = 0$, for type i may be able to kill type j, but type j may not be able to kill type i. Obviously, we examine only one matrix from each equivalence class, e.g. matrices

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \tag{13.2}$$

are equivalent and correspond to isomorphic graphs.

In the remainder of the chapter, we will refer to any equivalence class of matrices by giving one of its elements.

The number of interaction matrices increases rapidly with n : There are 15 nontrivial matrices with $n = 3$, 217 with $n = 4$, and 9608 with $n = 5$. These matrices can be generated using software *Nauty* (McKay and Piperno 2014).

All experiments with the same number of elements ($n = 3$, $n = 4$, or $n = 5$) of interaction matrices and the same type of field (hexagonal or square) are done always with the same initial field, randomly generated.

13.2.3 Quorum Sensing Model

The quorum sensing model extends the basic model introduced in Sect. 13.2.2. In this model, the interactions follow three steps, with prespecified threshold m:

1. Selection of a focal cell of type j.
2. Random selection of a neighbor of type i of the focal cell.
3. If $a_{ij} = 1$, then type i kills the focal cell of type j provided that there are at least m cells of type i in the neighborhood of the focal cell. The selected neighbor of type i cannot kill the focal cell if $a_{ij} = 0$, or if $a_{ij} = 1$ meanwhile there are less than m cells of type i in the neighborhood of the focal cell.

Figure 13.2 shows an example of the quorum sensing model, where the minimum quorum sensing threshold is set to $m = 2$. Consider the following interaction matrix:

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{13.3}$$

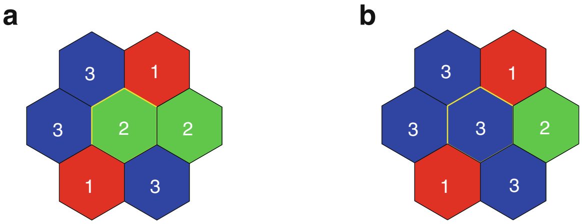

Fig. 13.2 An example of possible steps in the quorum sensing simulation. (**a**) A focal cell of type 2 and its neighbors. (**b**) Changing the focal cell type

A focal cell of type 2 is randomly selected. If a neighbor of type 1 or 2 is selected, nothing would happen as $a_{12} = a_{22} = 0$. However, if a neighbor of type 3 is selected and there are at least $m = 2$ type 3 neighbors of the focal cell, then the focal cell is replaced by a cell of type 3, because $a_{32} = 1$. The latter is illustrated in Fig. 13.2b.

13.3 Case Studies on Coexistence

In this section we examine a number of conditions which might influence coexistence of different types. By coexistence we mean that each type which was initially introduced is represented by at least one cell at the end of the simulation.

13.3.1 Case Study 1: The Order in Which Competing Cells Are Selected

This case study uses the model introduced in Sect. 13.2.2. Tests with three strains, four strains, and five strains will be carried out. The test field will consist of 8100 cells, equally distributed between the strains. Two different ways of how the focal cells are selected in each round, will be considered:

- Partially random: In each round all 8100 cells are selected, one by one, in random order. The basic interactions introduced in Sect. 13.2.2 are performed for each of them and the field is updated immediately after each interaction takes place for one selected focal cell. There are 12,000 rounds in total.
- Random: We randomly select 8100 cells and each time we perform the basic interactions introduced in Sect. 13.2.2. As in the previous case, the field is updated immediately after each interaction takes place. We can call these simulations "with random selection", because after each interaction a new focal cell is selected randomly from the entire field. In this case, some cells may be updated more than once, while some may not be updated at all.

For each interaction matrix of sizes 3×3 and 4×4, respectively, we ran 100 simulations (15×100 and 217×100, respectively). For each interaction matrix of size 5×5 we ran 10 simulations (9608×10). We only ran ten simulations in the latter case, because of the high computational time. Simulations were ran for two sets of conditions, partially random and random. As before, the same initial field is used for all runs and in each initial field all types are equally distributed.

We observe that a higher portion of interaction matrices leads to coexistence in the set of simulations with random selection than in the set of simulations with partially random selection. Why this happens is further explained:

In the set of simulations with partially random selection every cell interacts as a focal cell in every round. Such is not true for simulations with random selection. Therefore, in the simulation with partially random selection a larger part of the field is changed in each round. After a finite number of rounds, random selection simulations will still have some fraction of the initial field, because the cells in that fraction simply have not interacted with their neighbors yet.

The results are summarized in Fig. 13.3.

In order to test if there is a significant statistical difference between the results of the two simulation setups, we performed statistical non-parametric tests [Wilcoxon paired Signed-Rank tests (Bauer 1972; Hollander et al. 2013)] using the software package *R* on the simulation results. The Wilcoxon paired Signed-Rank test is suitable for comparing two matched samples. The test is an equivalent of the paired *t*-test that does not assume normality of the data (Oyeka and Ebuh 2012). We compared the simulation results in the case where focal cells are chosen with random selection to the simulation results in the case where focal cells are chosen with partially random selection. Three tests were performed, for the three strains, four strains, and five strains in the field. We found that the differences in simulation outcomes were not significant for the three strains (p-value $= 3 \cdot 10^{-1}$), but significant for the four strains (p-value $< 5 \cdot 10^{-3}$) as well as for the five strains (p-value $< 10^{-3}$). The results of the statistical tests confirm that a more random way of selecting the focal cell is correlated with a higher chance of coexistence.

13.3.2 Case Study 2: The Number of Neighbors

The simulations in this case study use the basic model described in Sect. 13.2.2. We examine how the coexistence evolves on two types of grids: one where each cell has six neighbors and one where each cell has eight neighbors, as introduced in Sect. 13.2.1. As before, interactions among three strains, four strains, and five strains are studied. For each interaction matrix of sizes 3×3 and 4×4, respectively, we ran 100 simulations and for each interaction matrix of size 5×5 we ran 10 simulations. The initial conditions are as before. Results are summarized in Fig. 13.4.

As in the previous case study, we ran Wilcoxon paired Signed-Rank tests (Bauer 1972; Hollander et al. 2013) to check the significance of the different simulation outcomes.

We found that the differences in simulation outcomes were not significant for the three strains (p-value $\approx 3 \cdot 10^{-1}$), but significant for the four strains (p-value

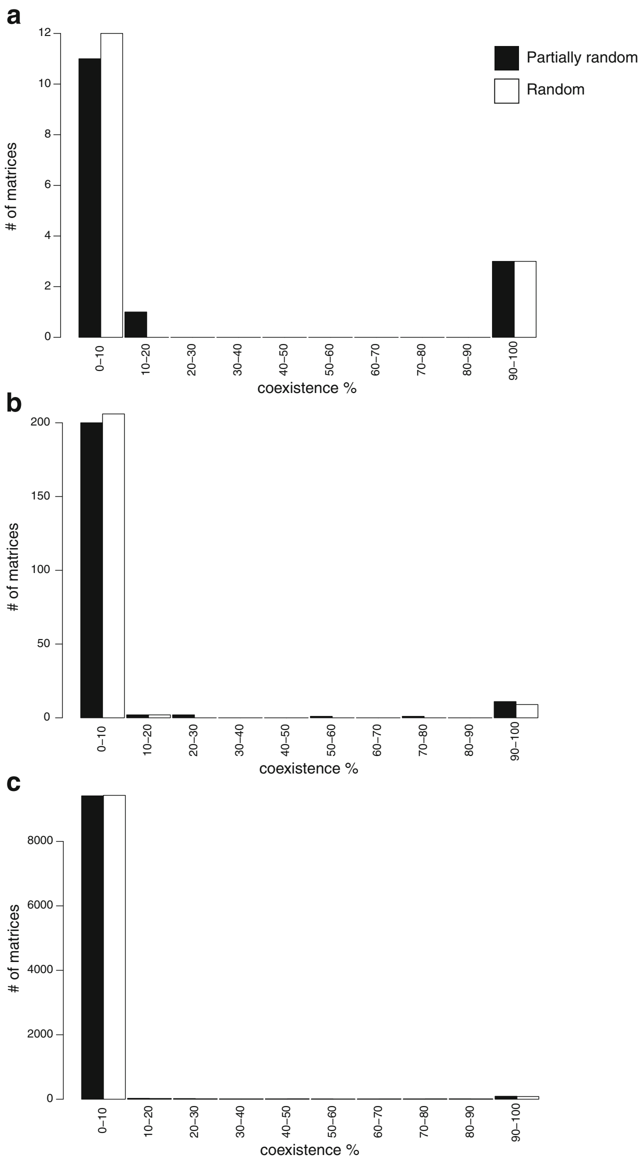

Fig. 13.3 Histograms showing simulation results. We ran two sets of simulations and each set included 100 simulations for each of the 3 × 3 (**a**) and 4 × 4 (**b**) matrices of interactions and 10 simulations for each of the 5 × 5 matrices (**c**). In one set of simulations cells interacted randomly and in the other set of simulations cells interacted in a partially random manner, as described in Sect. 13.3.1. We assigned each matrix a coexistence percentage, representing the percentage of simulations which lead to survival of all of the starting strains, out of 100 and 10 runs, respectively. For the two sets of simulations the histogram depicts the number of matrices that lead to various degrees of coexistence

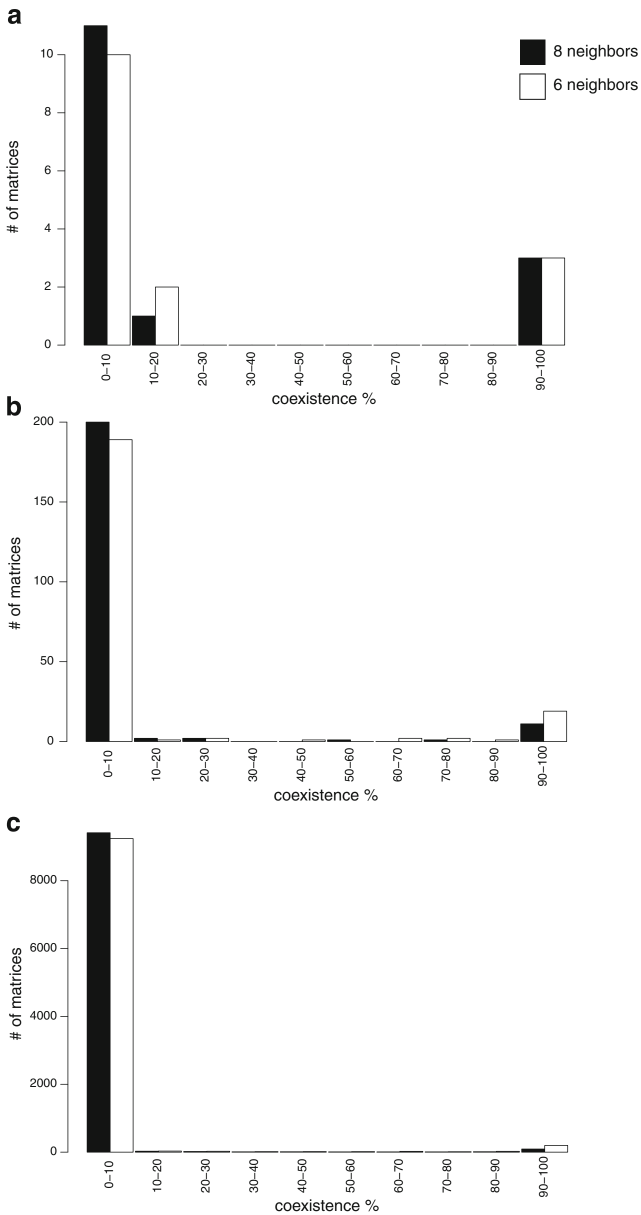

Fig. 13.4 Histograms showing simulation results. We ran two sets of simulations and each set included 100 simulations for each of the 3 × 3 and 4 × 4 interaction matrices and ten simulations for each of the 5 × 5 matrices. In one set of simulations cells had six neighbors and in one set of simulations cells had eight neighbors, as presented in Sect. 13.2.1. We assigned each matrix a coexistence percentage, representing the percentage of simulations which leads to 100 % survival of all of the starting strains, out of all runs. The histogram depicts the number of matrices that lead to different percentages of coexistence in two sets of simulations

$<5 \cdot 10^{-4}$), and for the five strains (p-value $<2.2 \cdot 10^{-16}$). The statistical tests confirm that the change from 6 to 8 neighbors for each focal cell decreases the fraction of interaction matrices that lead to coexistence. Intuitively, this seems natural as higher numbers of neighbors imply that it is more likely that a focal cell meets a predator and is killed.

13.3.3 Case Study 3: Basic Model vs. Quorum Sensing Model

We compared the simulation outcomes of the basic model (described in Sect. 13.2.2) with the simulation outcomes of the quorum sensing model (described in Sect. 13.2.3).

For each interaction matrix of sizes 3×3 and 4×4, respectively, we ran 100 simulations and for each interaction matrix of size 5×5 we ran 10 simulations, for each of two different setups: the first one using the basic model and the second using the quorum sensing model. In both setups discussed here, cells have six neighbors and the selection of the focal cell is done with random selection, as described in Sect. 13.3.1. The initial conditions are the same as before.

In the quorum sensing model we considered the situations where the quorum sensing threshold varies between 2 and 4, that is, the minimal number of neighbors of the focal cell belonging to the same strain required to kill should be at least 2, 3 or 4, respectively. Figure 13.5 shows our results.

Note that as m increases, the focal cell must have more neighbors of the same type to be killed, however also the number of types that meet this threshold goes down.

Our results confirm previous studies that showed that the quorum sensing plays a major role in promoting microbial coexistence. As seen in Fig. 13.5, the number of interaction matrices that lead to coexistence increases with the increase of quorum sensing threshold m.

Moreover, we showed that when the quorum sensing threshold is high (4 in our case) the topology of the interaction matrices does not influence coexistence. In the quorum sensing model 3, 4, or 5 strains can coexist regardless of the interaction matrices, as it can be seen in Fig. 13.5. Obviously, this happens because the threshold m is too high.

13.3.4 Case Study 4: Matrix Topology

In this case study we test how the topology of the interaction matrix affects the simulation results. We picked two metrics that describe the topology of the matrices of interactions: firstly, the density of the matrix and secondly, the number of cycles in the matrix. For 3×3 matrices we looked at cycles of sizes 2 and 3, for 4×4

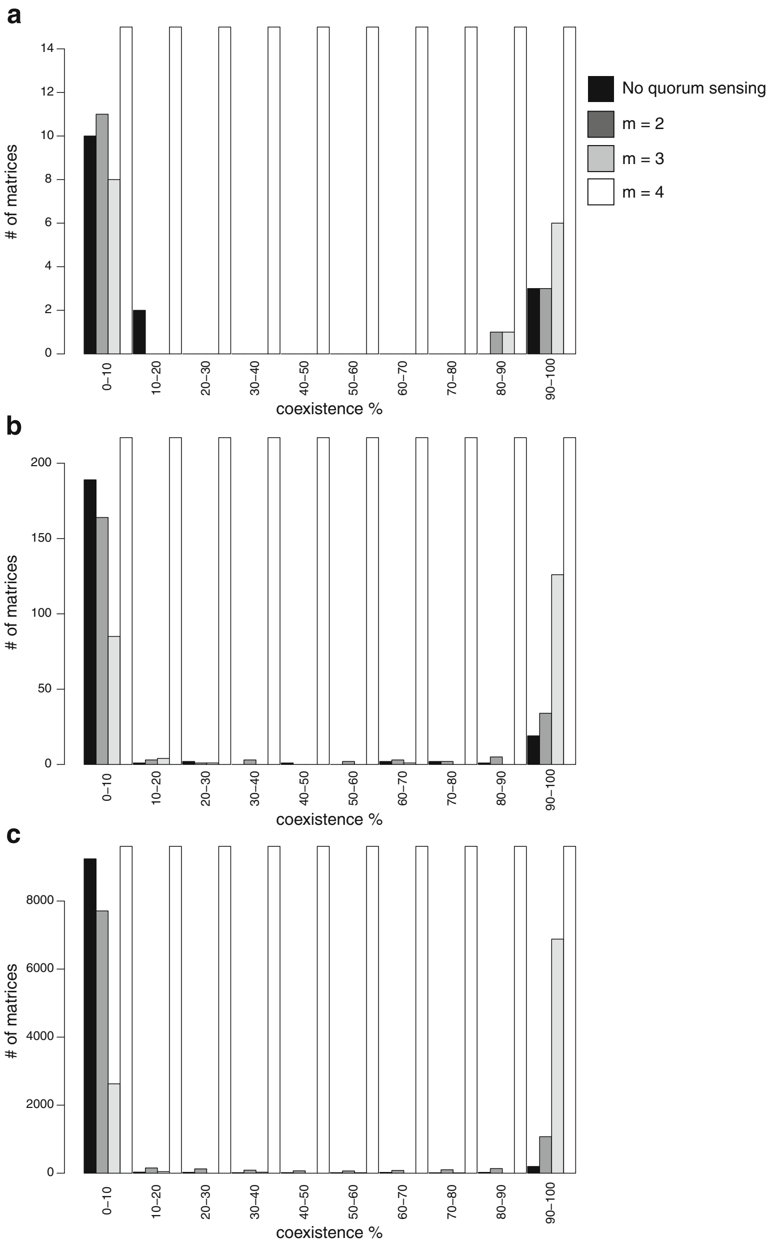

Fig. 13.5 Histograms showing simulation results of the quorum sensing model. We ran four sets of simulations [for $m = 1$ (the basic model), $m = 2$, $m = 3$ and $m = 4$ (the quorum sensing model)] and each set included 100 simulations for each of the 3×3 (**a**) and 4×4 (**b**) matrices of interactions and 10 simulations for each of the 5×5 matrices (**c**). We assigned each matrix a coexistence percentage, representing the percentage of simulations which lead to 100 % survival of all of the starting strains, out of 100 and 10 runs, respectively. For the fours sets of simulations the histogram depicts the number of matrices that lead to various degrees of coexistence

matrices we chose the cycles of sizes 2, 3, and 4, and for 5×5 matrices we chose the cycles of sizes 2, 3, 4 and 5. We report the results we found in the simulations where each cell has eight neighbors and focal cells are chosen with random selection, as explained in Sect. 13.3.1. We again ran 100 simulations for each of the 3×3 and 4×4 matrices and 10 simulations for each of the 5×5 matrices. The initial conditions are as before.

In order to test if the densities of the interaction matrices are correlated with the simulation results, we performed a set of Pearson correlation tests. The output of the correlation test, the correlation coefficient, can vary between -1 and 1, where -1 shows perfectly negatively correlated samples and 1 shows perfectly positively correlated samples. We found the following correlation coefficients: for the set of the 3×3 interaction matrices (densities and their respective simulation results) correlation coefficient 0.09; for the set of the 4×4 interaction matrices (densities and respective simulation results) correlation coefficient <0.001; for the set of the 5×5 interaction matrices (densities and respective simulation results) correlation coefficient <0.001.

Similarly, we tested whether there was any correlation between the total number of cycles in the interaction matrices and the simulation results. We found that the larger the interaction matrix the less correlated the cycles are with the simulation results (correlation coefficient 0.534 for the set of the 3×3 interaction matrices; correlation coefficient 0.185 for the set of the 4×4 interaction matrices; correlation coefficient <0.001 for the set of the 5×5 interaction matrices).

13.4 Remarks and Conclusions

We would like to stress that the mechanism in the simulations we ran is not the replicator dynamics frequently studied in evolutionary game theory. While for replicator dynamics random interactions are used to decide reproductive success of the interacting individuals, in our simulations the interactions are used to decide whether or not the focal individual will be killed and replaced by one of its neighbors. As such the matrices used in our models play a very different role from those used as fitness matrices in evolutionary games. It would be interesting to see whether the same results can be achieved using a replicator dynamics models.

As a second remark we wish to stress that in our simulations we observed something that one could call "Coexistence by small numbers". This means that quite frequently we observed that a specific species would not go extinct simply because of being completely surrounded and protected by harmless neighbors. Consider for example the following interaction matrix:

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{13.4}$$

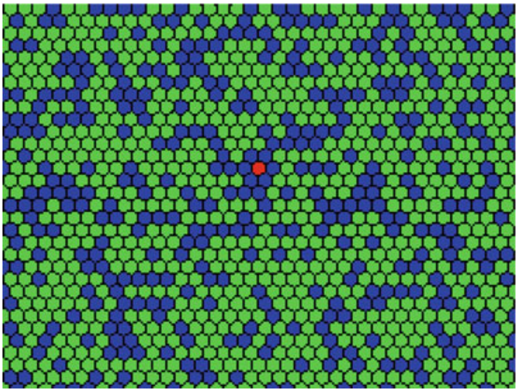

Fig. 13.6 An example of coexistence by small numbers. This is a snapshop of a hexagonal field at the last time step of a simulation where species follow the rules described in the interaction matrix (13.4)

The first species is killed by the second species (position (2,1) in the matrix), but it is immune to the third one (position (3,1) in the matrix). Few individuals of the first species can survive until the end time step (time step 12,000) in 100 simulations, by hiding inside islands formed by the third strain. Figure 13.6 shows the interaction field at the end time step of one of the simulations.

This simulation result is consistent with experimental studies where it was found that few strains dominate one habitat (McGill et al. 2007) and many others are present in low numbers. A new study demonstrated that the presence of the rare species is vital for the equilibrium of the community (Ren et al. 2014). Future simulations will investigate the mechanisms that lead to this experimental result. Moreover, we would like to be sure that this coexistence by small numbers is not a mere artifact of the tight and uniform neighborhood structures we examined. In the future we would like to examine models with more diverse neighborhood structures.

Our approach has three major limitations, which will be addressed in our future research. Firstly, the models presented assume that the interactions of n microbial species can be reduced to a binary matrix of size $n \times n$. Although this may seem very simplistic, the design of the model was motivated by previous experimental studies that characterized communities of coexisting microbes in a similar fashion (the results of those experimental studies were summarized in Table 13.1). Nevertheless, with the advance of experimental techniques, we expect that the future experimental characterizations of microbial communities will become more comprehensive and thus will allow us to improve our models, with the addition of new parameters such as, for example, the rate at which one type can kill another type. Secondly, the study investigated the potential coexistence of 3, 4 and 5 different microbial species. While this is informative, a future study should discuss coexistence of larger sets of strains. Thirdly, our models assume that all individual population members remain at fixed positions in space, while in reality the individuals move. We therefore wish to explore the effects of different mobility rules on coexistence and relate our models to other studies on this aspect (c.f. Reichenbach et al. 2007; Avelino et al. 2012).

Even though we are using simple models, our results showing that higher thresholds lead to higher rates of coexistence seem to be consistent with earlier findings (Diggle et al. 2007; Henke and Bassler 2004; Shapiro et al. 1998) stating that quorum sensing is a means of promoting coexistence among microbes.

Finally we would like to remark that all our simulations started from the assumption that all strains were initially well mixed and in equal proportions. Preliminary experiments suggest that having the strains clustered in patches dramatically influences some strains' ability to survive. In the future, we would like to approach this problem in a more systematic way.

References

Avelino PP, Bazeia D, Losano L, Menezes J (2012) Von Neummann's and related scaling laws in rock-paper-scissors-type games. Phys Rev E 86(3):031119

Bascompte J, Jordano P, Olesen JM (2006) Asymmetric coevolutionary networks facilitate biodiversity maintenance. Science 312(5772):431–433

Bauer DF (1972) Constructing confidence sets using rank statistics. J Am Stat Assoc 67(339):687–690

Chao L, Levin BR (1981) Structured habitats and the evolution of anticompetitor toxins in bacteria. Proc Natl Acad Sci 78(10):6324–6328

Consortium THMP (2012) Structure, function and diversity of the healthy human microbiome. Nature 486(7402):207–214

Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, Hirsch PR, Vogel TM (2011) Accessing the soil metagenome for studies of microbial diversity. Appl Environ Microbiol 77(4):1315–1324

Diggle SP, Griffin AS, Campbell GS, West SA (2007) Cooperation and conflict in quorum-sensing bacterial populations. Nature 450(7168):411–414

Dunne JA, Williams RJ, Martinez ND (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecol Lett 5(4):558–567

Dykhuizen DE (1998) Santa rosalia revisited: why are there so many species of bacteria? Antonie Van Leeuwenhoek 73(1):25–33

Funkhouser LJ, Bordenstein SR (2013) Mom knows best: the universality of maternal microbial transmission. PLoS Biol 11(8):e1001631

Henke JM, Bassler BL (2004) Bacterial social engagements. Trends Cell Biol 14(11):648–656

Hollander M, Wolfe DA, Chicken E (2013) Nonparametric statistical methods, vol 751. Wiley, New York

Hunt KM, Foster JA, Forney LJ, Schütte UM, Beck DL, Abdo Z, Fox LK, Williams JE, McGuire MK, McGuire MA (2011) Characterization of the diversity and temporal stability of bacterial communities in human milk. PLoS One 6(6):e21313

Kerr B, Riley MA, Feldman MW, Bohannan BJ (2002) Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. Nature 418(6894):171–174

Kinkel LL, Schlatter DC, Xiao K, Baines AD (2014) Sympatric inhibition and niche differentiation suggest alternative coevolutionary trajectories among streptomycetes. The ISME journal 8(2):249–256.

McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, Dornelas M, Enquist BJ, Green JL, He F et al (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecol Lett 10(10):995–1015

McKay BD, Piperno A (2014) Practical graph isomorphism, ii. J Symb Comput 60:94–112
Miller MB, Bassler BL (2001) Quorum sensing in bacteria. Annu Rev Microbiol 55(1):165–199
Montoya JM, Pimm SL, Solé RV (2006) Ecological networks and their fragility. Nature 442(7100):259–264
Oyeka ICA, Ebuh GU (2012) Modified wilcoxon signed-rank test. Open J Stat 2:172
Reichenbach T, Mobilia M, Frey E (2007) Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. Nature 448:1046–1049
Ren D, Madsen JS, Sørensen SJ, Burmølle M (2015) High prevalence of biofilm synergy among bacterial soil isolates in cocultures indicates bacterial interspecific cooperation. The ISME journal 9(1):81–89.
Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR et al (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science 341(6150):1241214
Riley MA, Gordon DM (1999) The ecological role of bacteriocins in bacterial competition. Trends Microbiol 7(3):129–133
Shapiro JA (1998) Thinking about bacterial populations as multicellular organisms. Annu Rev Microbiol 52(1):81–104
Thoemmes MS, Fergus DJ, Urban J, Trautwein M, Dunn RR (2014) Ubiquity and diversity of human-associated demodex mites. PloS One 9(8):e106265
Verhoef HA, Morin PJ (2010) Community ecology: processes, models, and applications. Oxford University Press, New York
Vetsigian K, Jajoo R, Kishony R (2011) Structure and evolution of streptomyces interaction networks in soil and in silico. PLoS Biol 9(10):e1001184
Waters CM, Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 21:319–346
Williams P, Winzer K, Chan WC, Camara M (2007) Look who's talking: communication and quorum sensing in the bacterial world. Philos Trans R Soc B 362(1483):1119–1134

Chapter 14
Computing α-Robust Equilibria in Two Integrated Assessment Models for Climate Change

Christopher Andrey, Olivier Bahn, and Alain Haurie

Abstract In this paper we show how to robustify the computation of equilibria in two integrated assessment models for climate change. Both models deal with the optimal timing of a transition to a 'clean' economy where a technology with low emissions but high energy cost can be used in the production process. The game represents the competition between industrialized and developing countries. A cost-benefit approach is implemented with an economic loss factor that represents the damages due to climate change. In the first model one assumes that both technologies, 'dirty' and 'clean' are available, but the economic loss factor is very uncertain. In the second model one assumes that the 'clean' technology is not yet available and some R&D investment must be made to get the technology breakthrough permitting its penetration. In this second model, formulated in continuous time, the jump rate of the controlled stochastic process describing the effect of R&D investment on the probability of breakthrough, is also considered as very uncertain. In both models we introduce a concept of α-robust equilibrium, where the robustification is achieved through the use of ambiguous probability distributions with a Kullback-Leibler divergence cost structure for the worst case choice by Nature.

At the time of this research the author "C. Andrey" was with ORDECSYS, Switzerland.

C. Andrey
ORDECSYS, Chêne-Bougeries, Switzerland
e-mail: christopher.andrey@gmail.com

O. Bahn (✉)
GERAD and Department of Decision Sciences, HEC Montréal, 3000, ch. Côte-Sainte-Catherine, Montréal, QC, Canada H3T 2A7
e-mail: olivier.bahn@gerad.ca

A. Haurie
ORDECSYS and University of Geneva, Geneva, Switzerland

GERAD, HEC Montréal, Montréal, QC, Canada
e-mail: ahaurie@gmail.com

© Springer International Publishing Switzerland 2016

F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_14

Keywords Integrated assessment modeling • Dynamic game • Robust equilibrium • Kullback-Leibler divergence • Piecewise deterministic differential game

MSC Codes: 91A10, 91A15, 91B62, 91B76

14.1 Introduction

Integrated assessment models (IAMs), like e.g. Bosetti et al. (2006), Gerlagh and van der Zwaan (2003), Leimbach et al. (2010), Manne et al. (1995), Manne and Richels (2005), Nordhaus and Boyer (2002), and Tol (2002a,b), combine a macroeconomic description of a production, investment, consumption, emission system for a group of countries with a schematized representation of the carbon cycle and resulting surface atmospheric temperature (SAT). An optimization approach, typically a deterministic optimal control one, is used to perform a cost-benefit analysis. The economic costs and benefits are evaluated through a utility function applied to consumption, whereas the cost due to climate change are evaluated through a damage function, linking the temperature change with a loss of economic output. These models are used to provide a "social value of carbon" and to give an indication of the optimal timing of abatement policies or of adaptation measures, as in Bahn et al. (2012). Several controversies have developed (see Nordhaus 2007; Tol 2008; Stern 2006) about the validity of the scenarios and carbon values derived from these models, motivated by the sensitivity of the models to the assumed value of key parameters, which remain very uncertain. This is particularly the case for the pure time preference rate, used to discount future utilities, and for the climate sensitivity parameter. In a recent paper Pindyck (2013) has concluded that these models could not be used to derive a correct social value for carbon. He has also advocated a cost effectiveness approach rather than a cost benefit one, to cope with the possibility of abrupt or catastrophic climate change. (Notice that Manne and Richels (1992) had already adopted a cost effectiveness approach with random long term limits on cumulative emissions to deal with an uncertain climate sensitivity.) In the same vein, Lempert and Collins (2007) have proposed a "robust approach" in the use of climate models. Babonneau et al. (2013) have implemented robust optimization technique to deal with the uncertainty in a meta-modeling approach to climate negotiations. In this game model cost effectiveness analysis was considered instead of the cost-benefit analysis usually applied when a Nash equilibrium was to be computed in an IAM as e.g. in Nordhaus and Yang (1996).

In this paper we propose the concept of α-robust equilibrium to deal directly with the uncertainty plaguing multi-country noncooperative IAMs. We implement the concept with a Kullback-Leibler divergence to measure the uncertainty deviation, in order to robustify equilibrium solutions in a multi-country cost-benefit model. For our demonstration we use two versions of an IAM already presented in Bahn and Haurie (2008) and Bahn et al. (2010), respectively. In these models one distinguishes between a 'carbon' and a 'low-carbon' economy. The first model is deterministic and we robustify the damage function, whereas the second model is stochastic and we robustify the controlled jump rate which describes the access to the clean technology as a breakthrough influenced by R&D investment.

The paper is organized as follows: in Sect. 14.2 we define the concept of α-robust equilibrium for a game described in its normal form, in Sect. 14.3 we adapt an approach recently proposed in Li et al. (2014) to robustify the damage function that drives the environmental game model proposed in Bahn and Haurie (2008); in Sect. 14.4 we show how to use the stochastic control method developed by Todorov (2009) to robustify the controlled jump rate in the stochastic game model proposed in Bahn et al. (2010); in Sect. 14.5 we conclude with an interpretation of the robust equilibrium solutions thus obtained. In this paper we focus on the technical game theoretic aspects of the approach. The significance of robust game theory models for climate change policy will be addressed in another paper.

14.2 Robust Equilibrium Concept

Consider an m-player game in normal form (Nash 1950) defined by the following data:

$$\begin{aligned} &\text{Players} && j = 1, \ldots, m \\ &\text{Payoffs} && \Psi_j(\mathbf{u}; \Theta), \quad j = 1, \ldots, m \\ &\text{Controls} && \mathbf{u} = (u_1, \ldots, u_m) \\ & && u_j \in U_j, \quad j = 1, \ldots, m. \end{aligned}$$

Θ is an uncertain parameter taking value in a space Ξ endowed with a pseudo-metric, or in the case where Θ is a measure of probability, a divergence,[1] denoted $\|\cdot\|$. This parameter has an a priori nominal value Θ^0.

Definition 1. An α-robust equilibrium $\mathbf{u}^\star$ satisfies:

$$\Phi_j(\mathbf{u}^\star) = \max_{u_j \in U_j} \Phi_j([u_j, \mathbf{u}^{\star -j}]), \quad j = 1, \ldots, m \tag{14.1}$$

where

$$\Phi_j(\mathbf{u}) = \Psi_j(\mathbf{u}; \Theta_j^\star) \tag{14.2}$$

and with

$$\Theta_j^\star = \operatorname*{argmin}_{\Theta \in \Xi} \{\Psi_j(\mathbf{u}; \Theta) + \alpha \|\Theta - \Theta^0\|\}, \quad j = 1, \ldots, m \tag{14.3}$$

[1] A divergence is a way to measure the distance between statistical distributions. Note that divergences need not satisfy the triangle inequality nor be symmetric.

We will consider in particular the case where Θ is a random variable and the pseudo-metric is given by the Kullback-Leibler divergence.

Remark 1. The intuition on the use of α-robust equilibrium can be given as follows: The parameter Θ is uncertain, so each player will consider a 'worst case', as if 'Nature' is a malevolent player trying to minimize the player's payoff. However to moderate the detrimental choice made by 'Nature' one assumes that there is a 'cost' for 'Nature' as the $\Theta_j^\star$ differs from the a priori nominal value Θ^0. Changing the weight α given to this cost in the choice of 'Nature' one controls the risk aversion for the players.

Remark 2. This definition of a robust equilibrium differs slightly from the one proposed in Aghassi and Bertsimas (2006), which is directly derived from robust optimization methods developed in Ben-Tal et al. (2009). Our definition is related to the approach of Hansen and Sargent (2008) for robust macroeconomic modeling.

14.3 A Model with Two Players, Two Technologies and Uncertain Damages

14.3.1 The Deterministic Model

The following indices and variables that enter in the description of the economic model are gathered in Table 14.1.

Table 14.1 Indices and variables

Symbol	Description
j	Index of each of the m regions
t	Running time (10 year periods)
$C(j,t)$	Total consumption in region j at time t, in trillions (10^{12}) of dollars
$I_i(j,t)$	Investment in capital i in region j at time t, in trillions of dollars
$K_i(j,t)$	Physical stock of productive capital i of region j at time t, in trillions of dollars
$L_i(j,t)$	Part of the (exogenously defined) labor force $L(j,t)$ of region j allocated at time t to economy i, in millions (10^6) of persons
$E_i(j,t)$	Yearly emissions of GHG (in Gt—10^9 tons—carbon equivalent) in the economy i of region j at time t
$Y_i(j,t)$	Economic output in the economy i of region j at time t, in trillions of dollars
$M(t)$	Atmospheric concentration of GHG at time t, in GtC equivalent
ELF(t)	Economic loss factor due to climate changes at time t, in %
WRG(j)	Discounted welfare of region j
W	Total discounted welfare

The economic description of each region j involves the following equations. First, a social planner is assumed to maximize social welfare (WRG), given by the sum over T 10-year periods of a discounted utility from available per capita consumption with the discount rate dr:

$$\mathrm{WRG}(j) = \sum_{t=0}^{T-1} 10\, dr(j,t)\, L(j,t) \ln\left[\frac{\mathrm{ELF}(t)\, C(j,t)}{L(j,t)}\right]. \tag{14.4}$$

Remark 3. Notice that we apply the economic loss factor ELF directly to consumption C and not, as it is usually done, to the output Y. This is motivated by the fact that we will soon consider this factor as a random variable, with ambiguous distribution. Applying the economic loss on consumption allows us to keep an open-loop formalism. Otherwise we would have been forced to introduce feedbacks as player strategies and develop a stochastic game formalism, with all the difficulties for implementing a numerical solution.

Total labor (L) is divided between labor allocated to the carbon economy (L_1) and labor allocated to the low-carbon economy (L_2):

$$L(j,t) = L_1(j,t) + L_2(j,t). \tag{14.5}$$

Capital stock (K_i) evolves through investment (I_i) and a depreciation rate δ_K as follows:

$$K_i(j,t+1) = 10\, I_i(j,t) + (1-\delta_K(j))^{10}\, K_i(j,t) \qquad i = 1,2. \tag{14.6}$$

Economic output (Y) occurs in the two economies according to an extended Cobb-Douglas production function in three inputs, capital (K), labor (L) and energy (which use is measured through emission level E):

$$\begin{aligned} Y(j,t) = {} & A_1(j,t)\, K_1(j,t)^{\alpha_1(j)}\, (\phi_1(j,t)\, E_1(j,t))^{\theta_1(j,t)}\, L_1(j,t)^{1-\alpha_1(j)-\theta_1(j,t)} \\ & + A_2(j,t)\, K_2(j,t)^{\alpha_2(j)}\, (\phi_2(j,t)\, E_2(j,t))^{\theta_2(j,t)}\, L_2(j,t)^{1-\alpha_2(j)-\theta_2(j,t)}, \end{aligned} \tag{14.7}$$

where A_i is the total factor productivity in the carbon (resp. low-carbon) economy when $i = 1$ (resp. $i = 2$), α_i is the elasticity of output with respect to capital K_i, ϕ_i is the energy conversion factor for emissions E_i and θ_i is the elasticity of output with respect to emissions E_i. Finally, economic output is used for consumption (C), investment (I) and the payment of energy costs:

$$Y(j,t) = C(j,t) + I_1(j,t) + I_2(j,t) + \pi_1(j,t)\, \phi_1(j,t)\, E_1(j,t) + \pi_2(j,t)\, \phi_2(j,t)\, E_2(j,t), \tag{14.8}$$

where π_i is the energy price in the carbon (resp. low-carbon) economy (when $i = 1$, resp. $i = 2$).

The simplified ‘climate module’ boils down here to computing the accumulation M of GHG in the atmosphere:

$$M(t+1) = 10\,\beta \sum_{j=1}^{m} (E_1(j,t) + E_2(j,t)) + (1-\delta_M)\,M(t) + \delta_M\,M_p, \tag{14.9}$$

where β is the marginal atmospheric retention rate (the fraction of emissions that remains in the atmosphere in the short run), δ_M is the natural atmospheric elimination rate (the rate of transfer from the atmosphere to the oceans) and M_p is the preindustrial level of atmospheric concentration (Tables 14.2 and 14.3).

Table 14.2 Nash solution vs α-robust Nash solution

	Nash		α-Robust Nash	
Period	Region-1	Region-2	Region-1	Region-2
2005	0	0	0	0
2015	0	0	0	0
2025	0	0	0	117
2035	0	0	0	191
2045	0	190	0	251
2055	0	290	0	308
2065	0	363	101	366
2075	0	427	195	489
2085	0	492	223	554
2095	0	554	223	613

Table 14.3 Pareto solution vs α-robust Pareto solution

	Pareto		α-Robust Pareto	
Period	Region-1	Region-2	Region-1	Region-2
2005	0	0	0	0
2015	63	0	62	0
2025	95	0	95	116
2035	120	153	119	190
2045	143	239	142	249
2055	168	303	166	305
2065	194	363	192	362
2075	223	424	220	422
2085	253	287	250	484
2095	283	550	280	548

14.3.2 Introducing Uncertainty in the Climate Change Damages

Damages due to climate change are summarized by the economic loss factor parameter ELF. These damages are uncertain. We shall assume that ELF is a function of a random parameter, with an ambiguous probability distribution for this random variable. More precisely:

Definition 2. ELF is defined by:

$$\mathrm{ELF}(t) = e^{-\bar{M}(t)\Gamma(t)}, \tag{14.10}$$

where, $\bar{M}(t) = M(t) - M(1990)$ and $\Gamma(t)$ is a random parameter that is a priori distributed according to an exponential law with probability density function (pdf) $\pi(\gamma) = \lambda e^{-\lambda\gamma}$.

Remark 4. In most IAMs, the ELF is defined as a function of temperature change. However, since this function is highly uncertain and since the temperature change is induced by the change in concentration, it makes sense to consider that *ELF* depends also in a highly uncertain manner on the evolution of *GHG* concentration, $M(t)$.

Γ can be interpreted as a random intensity of damages. For a given concentration $\bar{M}$, when Γ increases, the ELF parameter decreases, so more losses are incurred. The expected reward at time t is given by:

$$L(j,t)\int_0^{\infty} \ln\left[e^{-\bar{M}(t)\gamma}\frac{C(j,t)}{L(j,t)}\right]\lambda e^{-\lambda\gamma}\,d\gamma = L(j,t)\ln\left[e^{-\frac{\bar{M}(t)}{\lambda}}\frac{C(j,t)}{L(j,t)}\right]. \tag{14.11}$$

At each time t, the true distribution of the damage parameter $\Gamma(t)$ is ambiguous. Let $m(\gamma) = \frac{\hat{\pi}(\gamma)}{\pi(\gamma)}$ be the likelihood ratio of the worst distribution w.r.t. the a priori one. The K-L divergence between $\hat{\pi}(\gamma)$ and $\pi(\gamma)$, or relative entropy, is given by:

$$\Xi(\hat{\pi}(\gamma), \pi(\gamma)) = \int_{\Gamma} [m(\gamma)\ln[m(\gamma)]]\,\pi(\gamma)\,d\gamma, \tag{14.12}$$

where the integral is over the support of the random variable γ.

In the characterization of an α-robust equilibrium, we will assume that each player will consider the function $m(\cdot,t)$ as a control used by a malevolent player having the following criterion:

$$J(j) = \sum_{t=0}^{T-1} 10\,dr(j,t)\min_{m(\cdot,t)}\left\{L(j,t)\int_0^{\infty}\left(m(\gamma,t)\,\ln\left[e^{-\bar{M}(t)\gamma}\frac{C(j,t)}{L(j,t)}\right]\right.\right.$$
$$\left.\left.+\alpha\, m(\gamma,t)\ln[m(\gamma,t)]\right)\pi(\gamma)\,d\gamma\right\}, \tag{14.13}$$

where the likelihood ratio control of nature is subject to the constraints:

$$0 \leq m(\gamma(t), t) \tag{14.14}$$

$$1 = \int m(\gamma(t), t)\, \pi(\gamma(t))\, d\gamma(t). \tag{14.15}$$

Writing the necessary optimality conditions (NOCs) for the minimizing $m(\cdot, t)$, we obtain:

$$0 = \int \ln[e^{-\bar{M}(t)\gamma(t)} \frac{C(j,t)}{L(j,t)}] + (\alpha(\ln[m(\gamma(t)] + 1) + \kappa)\pi(\gamma(t))\, d\gamma(t), \tag{14.16}$$

where κ is a Lagrange multiplier associated with constraint (14.15). This yields to:

$$m^{\star}(\gamma(t)) = \frac{\exp\left(-\frac{\ln[e^{-\bar{M}(t)\gamma(t)} \frac{C(j,t)}{L(j,t)}]}{\alpha}\right)}{\int \exp\left(-\frac{\ln[e^{-\bar{M}(t)\gamma(t)} \frac{C(j,t)}{L(j,t)}]}{\alpha}\right) \pi(\gamma(t))\, d\gamma(t)} \tag{14.17}$$

as a solution to the NOCs for a minimizing likelihood ratio.

Since:

$$\exp\left(-\frac{\ln[e^{-\bar{M}(t)\gamma(t)} \frac{C(j,t)}{L(j,t)}]}{\alpha}\right) = \left(e^{-\bar{M}(t)\gamma(t)} \frac{C(j,t)}{L(j,t)}\right)^{\frac{-1}{\alpha}}, \tag{14.18}$$

the second factor cancels when expressing $m^{\star}(t)$, which is thus given by:

$$m^{\star}(\gamma, t) = \frac{e^{\frac{\bar{M}(t)}{\alpha}\gamma}}{\int_0^{\infty} e^{(\frac{\bar{M}(t)}{\alpha} - \lambda)\gamma} \lambda\, d\gamma}. \tag{14.19}$$

We finally get:

$$\pi^{\star}(\gamma(t)) = \left(\lambda - \frac{\bar{M}(t)}{\alpha}\right) e^{-\left(\lambda - \frac{\bar{M}(t)\, L(j,t)}{\alpha}\right)\gamma(t)}. \tag{14.20}$$

So the worst distribution is also an exponential law with intensity $\lambda - \frac{\bar{M}(t)}{\alpha}$, which is well defined provided the parameter α has been be chosen so that the expression $\alpha\lambda - \bar{M}(t)$ remains positive, that is:

$$\alpha \geq \frac{\bar{M}(t)}{\lambda}. \tag{14.21}$$

Remark 5. When $\alpha \to \infty$, we keep the a priori distribution. When α decreases, the intensity tends also to decrease and the expected value for the intensity of damages Γ tends to increase.

14.3.3 Robustified Payoffs and α-Robust Equilibrium

For a given α, we have characterized the worst distribution as an exponential law with intensity $\lambda - \frac{\bar{M}(t)}{\alpha}$. Therefore, in accordance with Definition 1, the α-robust equilibrium is the equilibrium computed with the robustified payoffs:

$$\int_0^\infty \ln\left[e^{-\bar{M}(t)\gamma}\frac{C(j,t)}{L(j,t)}\right]\pi^*(\gamma)\,d\gamma = \\ \int_0^\infty \left(\ln[e^{-\bar{M}(t)\gamma}] + \ln\left[\frac{C(j,t)}{L(j,t)}\right]\right)\left(\lambda - \frac{\bar{M}(t)}{\alpha}\right)e^{-(\lambda-\frac{\bar{M}(t)}{\alpha})\gamma}\,d\gamma. \tag{14.22}$$

The first term is given by:

$$\int_0^\infty -\bar{M}(t)\gamma\left(\lambda - \frac{\bar{M}(t)}{\alpha}\right)e^{-(\lambda-\frac{\bar{M}(t)}{\alpha})\gamma}\,d\gamma = -\frac{\alpha\,\bar{M}(t)}{\alpha\lambda - \bar{M}(t)}, \tag{14.23}$$

while the second term is $\ln[\frac{C(j,t)}{L(j,t)}]$. So, the robust reward function is defined by

$$\ln\left[e^{-\frac{\alpha\bar{M}(t)}{\alpha\lambda-\bar{M}(t)}}\right] + \ln\left[\frac{C(j,t)}{L(j,t)}\right] = \ln\left[e^{-\frac{\alpha\bar{M}(t)}{\alpha\lambda-\bar{M}(t)}}\frac{C(j,t)}{L(j,t)}\right] \tag{14.24}$$

The expected robust reward for Player j is then given by:

$$J(j,\sigma;s^0) = \sum_{t=0}^{T-1} 10\,dr(j,t)\,L(j,t)\,\ln\left[e^{-\frac{\alpha\bar{M}(t)}{\alpha\lambda-\bar{M}(t)}}\frac{C(j,t)}{L(j,t)}\right], \tag{14.25}$$

whereas the minimized cost for nature is given by:

$$J(j,\sigma;s^0) = \sum_{t=0}^{T-1} 10\,dr(j,t)\,L(j,t)\left\{\ln\left[e^{-\frac{\alpha\bar{M}(t)}{\alpha\lambda-\bar{M}(t)}}\frac{C(j,t)}{L(j,t)}\right]\right. \\ \left. +\alpha\int_0^\infty m^\star(\gamma,t)\ln[m^\star(\gamma,t)]\pi(\gamma)\,d\gamma\right\},$$

with

$$\alpha \int_0^\infty m^\star(\gamma,t)\ln[m^\star(\gamma,t)]\pi(\gamma)\,d\gamma = \alpha\left(\ln\left[1-\frac{\bar{M}(t)\,L(j,t)}{\alpha\,\lambda}\right]+\frac{\bar{M}(t)}{\alpha\lambda-\bar{M}(t)}\right). \tag{14.26}$$

14.3.4 Numerical Illustration

For this numerical illustration we use the same parameter values as in Bahn and Haurie (2008) to show the effect of robustification on the equilibrium solution. The world is divided into two regions: the first one (region $j = 1$) representing the developed countries, the second region ($j = 2$) representing the developing countries.

14.3.4.1 Numerical Results—Nash

We first compare the Nash equilibrium solution and the α-robust equilibrium, looking more precisely at the accumulation of capital in the low-carbon economy (variable K_2).

Remark 6. We notice an important effect of the robustification on the equilibrium solution. Capital accumulation in the low-carbon economy occurs much earlier for both players.

14.3.4.2 Numerical Results—Pareto

Next, we do a similar comparison in the context of a Pareto equilibrium solution, obtained when optimizing a weighted sum of the regional welfares with equal weight given to each player. The next table reports again on the accumulation of capital in the low-carbon economy.

Remark 7. We notice a much lower effect of the robustification on the Pareto solution. Capital accumulation in the low-carbon economy is not much affected by the robustification. This is probably due to the fact that the Pareto solution has already reduced notably atmospheric GHG concentration and thus climate change damages.

14.4 Optimal R&D Investment in a Stochastic Continuous Time Model

In this section we introduce robustification in the stochastic game model proposed in Bahn et al. (2010). The key uncertainty considered in this model is the controlled jump rate, which describes the probability of accessing to the clean, more efficient technology. We use again K-L divergence in a formulation of the nature minimization problem along the line of the general approach proposed by Todorov (2009) to formulate and solve controlled Markov chains. The α-robust stochastic equilibrium is formulated and a numerical illustration is provided.

14.4.1 A Continuous Time Framework

The dynamics is now represented by differential state equations:

$$\dot{K}_i(j,t) = I_i(j,t) - \delta_i K_i(j,t), \quad i = 1,2,\ j = 1,\dots,m,$$
$$\dot{M} = \sum_{j=1}^{m} (E_1(j,t) + E_2(j,t)) - \delta_M \,(M(t) - M_p),$$

and the payoffs are given by random integrals

$$J(j) = \mathbb{E}\left[\int_0^\infty e^{-\rho(j)t} L(j,t) \ln\left[e^{-\frac{\bar{M}(t)}{\lambda}} \frac{C(j,t)}{L(j,t)}\right] dt|\mathbf{K}(0), M(0)\right], \quad j = 1,\dots,m,$$

where the expectations are taken with respect to the probability measure associated with the stochastic jump process describing access to the clean technology.

Here again we consider a two-player game, i.e. we take $m = 2$.

14.4.2 Technological Breakthrough Dynamics

We assume that the technological breakthrough is driven by a stochastic jump process with an uncertain jump rate.

Variable: To describe the access to the cleaner productive system a binary variable $\xi \in \{0,1\}$ is introduced which indicates if the clean technology is available ($\xi = 1$) or not yet ($\xi = 0$).

Breakthrough dynamics: The initial value $\xi(0) = 0$ indicates that there is no access to the clean capital at the initial time. The switch to the value 1 occurs at a random time which is controlled through the global accumulation of R&D capital $\tilde{K}_2$ where:

$$\tilde{K}_2(t) = \sum_{j=1}^{2} K_2(j,t). \tag{14.27}$$

However this jump rate is uncertain. More precisely we assume that the elementary probability of a breakthrough is given by

$$P[\xi(t+dt) = 1|\xi(t) = 0, \tilde{K}_2(t)] = \Theta(t)\,dt + o(dt),$$

where

- $\Theta(t)$ is a positive jump rate control that can be chosen by Nature with a K-L running cost

$$\alpha\left(\Theta(t)\ln\left[\frac{\Theta(t)}{q_b(t,\tilde{K}_2(t))}\right] + q_b(t,\tilde{K}_2(t)) - \Theta(t)\right). \tag{14.28}$$

- where $q_b(t,\tilde{K}_2(t)) = \omega_b + \upsilon_b\tilde{K}_2(t)$ is the a priori nominal jump rate function, where the parameters take the following values

  ω_b : initial probability rate of discovery; $\omega_b = 0.05$;
  υ_b : slope w.r.t $\tilde{K}_2(t)$ of the probability rate of discovery; $\upsilon_b = 0.0019$.

Remark 8. One can easily verify that the running cost defined in (14.28) is always ≥ 0 and attains its minimum, equal to 0, when $\Theta(t) = q_b(t,\tilde{K}_2(t))$. It is directly related to the K-L divergence for the elementary jump probabilities.

- Assume that the breakthrough jump occurs at time τ. From time τ onwards, the carbon-free technology is available.
- At time τ, for a given state $s = (\mathbf{K}, M)$, a value function $V(j,\tau,\bar{x})$ is defined as the current-valued payoff to Player j in the equilibrium solution to the robust open-loop dynamic game defined as previously, but in a continuous time setting. Here $\bar{x}$ refers to the initial values of all state variables, K_1, K_2 for all the players and M.

  The control problem that we assume to be solved by nature, is written as follows:

$$\min_{\Theta(\cdot)} \mathbb{E}\left[\int_0^{\tau} \alpha\left(\Theta(t)\ln\left[\frac{\Theta(t)}{q_b(t,\tilde{K}_2(t))}\right] + q_b(t,\tilde{K}_2(t)) - \Theta(t)\right) dt \right.$$
$$\left. + V(j,\tau,\bar{x}(\tau)) - V^0(j,\tau,\bar{x}(\tau))\right] \tag{14.29}$$

s.t. state dynamics and jump stochastics. Here $V(j,\tau,\bar{x}(\tau))$ is the current valued expected cost to go (or player-j's payoff) from time τ onward, when the new technology is available (the jump has occurred). The function $V^0(j,\tau,\bar{x}(\tau))$ is current valued expected cost to go (or the player payoff) from time τ onward, if the new technology is never available (no jump can occur).

Remark 9. This performance criterion used by Nature expresses the fact that the jump rate, used by the players in the computation of an equilibrium, will be determined in such a way that it reduces the payoff advantage given by an access to the new technology, while not diverging too much from the nominal jump rate function. This criterion for Nature is slightly different from the one used in Definition 1 when introducing the α-equilibrium concept; however the general idea is similar, Nature seeks to reduce the gains of the players.

14.4.3 Tenet of Transition and Optimal Control of Nature

The control problem that we assume to be solved by Nature is defined in (14.29). Expliciting the probability density of the random jump time, and integrating by parts leads to the deterministic equivalent control problem

$$\min_{\Theta(\cdot)} \int_0^\infty e^{-\int_0^t \Theta(s)\,ds} \Big\{ \Theta(t)(V(j,t,\bar{x}(t)) - V^0(j,t,\bar{x}(t))) + \alpha \left(\Theta(t) \ln\left[\frac{\Theta(t)}{q_b(t,\tilde{K}_2(t))}\right] + q_b(t,\tilde{K}_2(t)) - \Theta(t) \right) \Big\}\, dt. \tag{14.30}$$

Introduce the so-called desirability function $z(j,t,\bar{x}(t)) = e^{-\frac{V(j,t,\bar{x}(t)) - V^0(j,t,\bar{x}(t))}{\alpha}}$. Then Eq. (14.30) becomes

$$\min_{\Theta(\cdot)} \int_0^\infty e^{-\int_0^t \Theta(s)\,ds} \alpha \Big(\Theta(t) \Big(-\ln[z(j,t,\bar{x}(t))] + \ln[\frac{\Theta(t)}{q_b(t,\tilde{K}_2(t))}] \Big) + q_b(t,\tilde{K}_2(t)) - \Theta(t) \Big)\, dt \tag{14.31}$$

$$= \min_{\Theta(\cdot)} \int_0^\infty e^{-\int_0^t \Theta(s)\,ds} \alpha \Big(\Theta(t) \ln[\frac{\Theta(t)}{q_b(t,\tilde{K}_2(t)) z(j,t,\bar{x}(t))}] + q_b(t,\tilde{K}_2(t)) - \Theta(t) \Big)\, dt. \tag{14.32}$$

As noticed in Remark 8 above, the minimum of the expression

$$\Theta(t) \ln[\frac{\Theta(t)}{q_b(t,\tilde{K}_2(t)) z(j,t,\bar{x}(t))}] + q_b(t,\tilde{K}_2(t)) z(j,t,\bar{x}(t)) - \Theta(t)$$

is equal to 0 when

$$\Theta(t) = q_b(t,\tilde{K}_2(t)) z(j,t,\bar{x}(t)) = q_b(t,\tilde{K}_2(t))\, e^{-\frac{V(j,t,\bar{x}(t)) - V^0(j,t,\bar{x}(t))}{\alpha}}. \tag{14.33}$$

As the rest of the running cost does not depend on $\Theta(t)$ this defines the optimal control chosen by Nature.

Remark 10. Notice that the robustified jump rate is the a priori nominal jump rate multiplied by a factor which depends on the cost-to-go or value function computed for the after jump game equilibrium.

- When $\alpha \to \infty$ then $\Theta(t) \to q_b(t, \tilde{K}_2(t))$
- When $\alpha \to 0$ then $\Theta(t) \to 0$.

14.4.4 Computing an α-Robust Equilibrium

1. Solve, for an ensemble of initial states (space filling set) and times, the robustified open-loop DG, when the new technology is available;
2. Adjust an analytical form to the set of cost-to-go and obtain an approximation of $V(j, t, \bar{x}(t))$;
3. Then solve the open-loop DG defined by the modified payoffs

$$V^0(j, \bar{x}^0) = \max_{u(j,\cdot)} \int_0^{\infty} e^{-(\rho(j)t + \int_0^t q_b(t, \tilde{K}_2(s)) \exp[-\frac{V(j,s,\bar{x}(t)) - V^0(j,s,\bar{x}(t))}{\alpha}] ds)}$$
$$\{L(j,t) \ln[e^{-\bar{M}(t)\gamma(t)} \frac{C(j,t)}{L(j,t)}] + q_b(t, \tilde{K}_2(t))$$
$$\exp[-\frac{V(j,t,\bar{x}(t)) - V^0(j,t,\bar{x}(t))}{\alpha}] V(j,t,\bar{x}(t))\} dt \qquad (14.34)$$

   s.t. the state equations.
4. Look at how the solution varies when α decreases.

Remark 11. By adapting the results of previous section to a continuous time setting, it is possible to combine both a robustification of *ELF* and a robustification of the technology breakthrough rate.

14.4.5 Numerical Illustration

14.4.5.1 Parameters

We have run the model with the following parameter values:

Energy prices in carbon economy: $\pi_1(1) = 0.35 \quad \pi_1(2) = 0.3$;
Energy prices in carbon-free economy: $\pi_2(1) = 0.75 \quad \pi_1(2) = 0.8$;
Rate of social time preference: 3 %;

Total factor productivity: Initial values $A_i(j,t) \equiv 0.0302$, initial growth rate $gA_i(j,t) \equiv 0.1$, rate of decrease of growth rate $dgA_i(j,t) \equiv 0.005$
Capital elasticity in production function: $\alpha_i(1) \equiv 0.3$, $\alpha_i(2) \equiv 0.35$;
Annual capital depreciation rate: 10N %;
World population levels (in million): $L(1,0) = 2205$, $L(2,0) = 4205$, initial growth rate $gL(1,0) = 0.08$, $gL(2,0) = 0.08$, rate of decrease of growth rate $dgL(1,t) = 0.3$, $dgL(2,t) = 0.3$
Energy efficiency in carbon economy: $\phi_1(1,0) = 1.1$ $\phi_1(2,0) = 0.9$, initial growth rate $g\phi_1(1,0) = 0.15$, $gL(2,0) = 0.15$, rate of decrease of growth rate $dgL(1,t) = 0.2$, $dgL(2,t) = 0.2$;
Energy efficiency in carbon-free economy: $\phi_2(1,0) = 3$ $\phi_1(2,0) = 2.5$, initial growth rate $g\phi_1(1,0) = 0.15$, $gL(2,0) = 0.15$, rate of decrease of growth rate $dgL(1,t) = 0.2$, $dgL(2,t) = 0.2$;
Energy elasticity in production function: $\theta_i(1,0) \equiv 0.05$, $\theta_i(2,0) \equiv 0.05$, initial growth rate $g\theta_i(1,0) = -0.012$, $g\theta_i = 0.012$, rate of decrease of growth rate $dg\theta_i(1,0) = -0.008$, $dg\theta_i = 0.008$, $i = 1,2$;
Carbon concentration in 1990: 750;
K-L parameter for damage function: $\alpha_1 = 1$;
Expected value for random damage function parameter: $\lambda = 1250$;
Initial values for capital stocks: $K_1(j,0) \equiv 48.65$, $K_2(j,0) \equiv 0$.

14.4.5.2 Effect of Robustification

We show below the accumulation paths for "dirty" $K_1(j,t)$ and "clean" $K_2(j,t)$ capital stocks when one varies the K-L coefficient α. Because of the orders of magnitude of the value functions and the K-L divergence cost, the α values must be quite large.

14.4.5.3 Interpretation

A full exploitation of these numerical simulations will be the object of another paper. The present one is dedicated to the description of the method. In the case under study we have two regions in the world, with the same initial stock of "dirty" capital, i.e. a carbon economy. However the population size is much larger in region 2, which corresponds grossly to developing countries, whereas region 1 corresponds more or less to industrialized countries. The clean capital stock is essentially used as an R&D asset, which influences the probability of having access to a carbon-free economy. Recall that the jump rate for a technological breakthrough is $q_b(t,\tilde{K}_2(t)) = \omega_b + \upsilon_b\tilde{K}_2(t)$. In this numerical illustration we have chosen $\omega_b = 0$ and $\upsilon_b = 0.1$.

We see in Table 14.4 and Fig. 14.1, that for a "low" value $\alpha = 900{,}000$, the pattern of investment in the R&D asset is higher than for the cases with "high" value $\alpha = 9{,}000{,}000$ or $\alpha > 9{,}000{,}000$. For these high values of α the solution tends to stabilize with a capital stock $K_2(2,t)$ staying around 6 trillion \$ for the developing

Table 14.4 $\alpha = 900{,}000$

	2005	2015	2025	2035	2045	2055	2065	2075	2085	2095
$K_1(\cdot, , \cdot)$										
Country 1	48.65	51.80	64.65	79.24	95.48	113.41	131.98	153.79	139.67	131.94
Country 2	48.65	68.20	78.76	97.14	119.92	146.40	176.26	208.80	243.37	166.53
$K_2(\cdot, , \cdot)$										
Country 1	0.00	8.04	0.51	0.27	0.22	0.14	0.15	0.12	0.10	0.00
Country 2	0.00	0.41	8.16	9.04	9.34	9.58	9.43	9.14	8.77	7.42

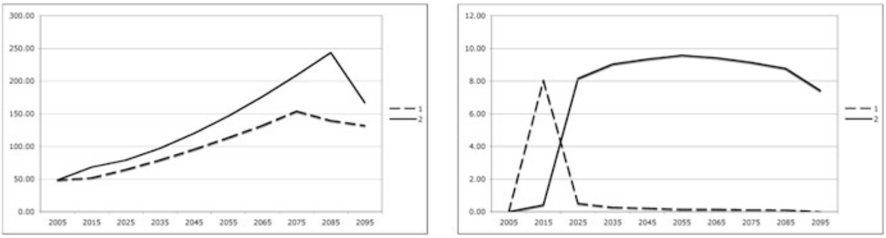

Fig. 14.1 K_1 and $K_2 - \alpha = 900{,}000$

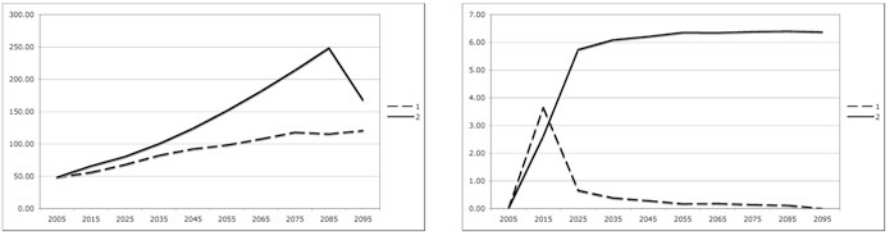

Fig. 14.2 K_1 and $K_2 - \alpha = 9{,}000{,}000$

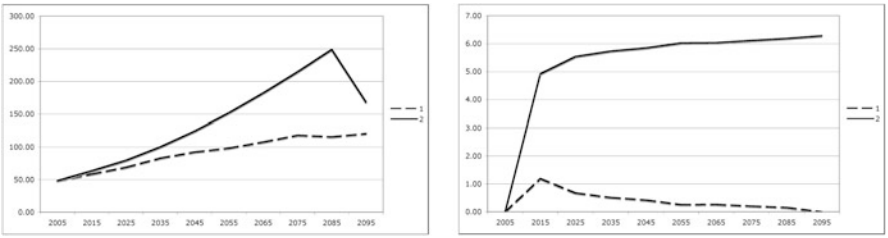

Fig. 14.3 K_1 and $K_2 - \alpha = 90{,}000{,}000$

countries and a much lower R&D capital stock $K_2(1, t)$ for industrialized countries, which, after an initial surge at 1.19 trillion \$ is left declining (Figs. 14.2 and 14.3).

When $\alpha = 900{,}000$, i.e. for a low weight given to the K-L divergence cost, the possible perturbation due to nature choice of a detrimental jump rate is more important. Therefore in the α-robust equilibrium solution the R&D capital stock

Table 14.5 $\alpha = 9{,}000{,}000$

	2005	2015	2025	2035	2045	2055	2065	2075	2085	2095
$K_1(\cdot,,\cdot)$										
Country 1	48.65	56.19	67.96	82.45	92.36	98.37	107.70	118.18	115.60	120.84
Country 2	48.65	66.00	80.41	100.34	124.19	151.40	181.58	214.08	248.35	167.98
$K_2(\cdot,,\cdot)$										
Country 1	0.00	3.65	0.66	0.39	0.29	0.18	0.19	0.15	0.12	0.00
Country 2	0.00	2.61	5.74	6.09	6.22	6.36	6.35	6.39	6.41	6.38

Table 14.6 $\alpha > 90{,}000{,}000$

	2005	2015	2025	2035	2045	2055	2065	2075	2085	2095
$K_1(\cdot,,\cdot)$										
Country 1	48.65	58.64	69.20	83.13	92.52	98.57	107.84	118.08	115.72	120.65
Country 2	48.65	63.69	79.83	100.23	124.35	151.72	181.99	214.51	248.76	168.09
$K_2(\cdot,,\cdot)$										
Country 1	0.00	1.19	0.68	0.52	0.42	0.26	0.27	0.20	0.16	0.00
Country 2	0.00	4.92	5.54	5.74	5.85	6.03	6.04	6.12	6.19	6.29

$K_2(2,t)$ stays around 9 trillion \$ for the developing countries and is still much lower for industrialized countries, since $K_2(1,t)$, after an initial surge at 8 trillion \$ is left declining (Tables 14.5 and 14.6).

14.5 Conclusion

In this paper we have shown how to introduce robustness in the computation of Nash equilibrium solution for multi-country IAMs. Using the K-L divergence as a cost to moderate the detrimental choice of a probability distribution by Nature, we have been able to obtain closed form solutions for the robust payoffs of the players in two versions of a multi-country IAM implementing cost-benefit analysis. The numerical illustration provided shows that the robustification of the game has an important influence on the solution. We advocate for the use of this robust game theory approach or the more direct one developed in Aghassi and Bertsimas (2006) in future analyses performed on multi-country IAMs. The clear advantage of the α-robust equilibrium with K-L divergence is that we can obtain closed form solution for the minimization performed by Nature.

Acknowledgements This research has been supported by the Natural Sciences and Engineering Research Council of Canada (O. Bahn) and by the Qatar National Research Fund under Grant Agreement no 6-1035-5126 (A. Haurie).

References

Aghassi M, Bertsimas D (2006) Robust game theory. Math Program Ser B 107:231–273

Babonneau F, Haurie A, Vielle M (2013) A robust meta-game for climate negotiations. Comput Manag Sci. 10(4):299–329. Available on the web

Bahn O, Haurie A (2008) A class of games with coupled constraints to model international GHG emission agreements. Int Game Theory Rev 10:337–362

Bahn O, Haurie A, Malhamé R (2010) A stochastic control/game approach to the optimal timing of climate policies. In: Filar JA, Haurie A (eds) Uncertainty and environmental decision making. Springer, Berlin, pp 211–237

Bahn M, Chesney A, Gheyssens J (2012) The effect of proactive adaptation on green investment. Environ Sci Policy 18:9–24

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization. Princeton University Press, Princeton

Bosetti V, Carraro C, Galeotti M, Massetti E, Tavoni M (2006) WITCH: a world induced technical change hybrid model. Energy J, 13–38 (Special Issue on Hybrid Modelling of Energy Environment Policies: Reconciling Bottom-up and Top-down)

Gerlagh R, van der Zwaan BCC (2003) Gross world product and consumption in a global warming model with endogenous technological change. Resour Energy Econ 25:35–57

Hansen L, Sargent T (2008): Robustness. Princeton University Press, Princeton

Leimbach M, Bauer N, Baumstark L, Edenhofer O (2010) Mitigation costs in a globalized world: climate policy analysis with REMIND-R. Environ Model Assess 15:155–173

Lempert R, Collins MT (2007) Managing the risk of uncertain threshhold responses: comparison of robust, optimum, and precautionary approaches. Risk Anal 27(4):1009–1026

Li X, Narajabad N, Borghan, Temzelides, Ted P (2014) Robust Dynamic Optimal Taxation and Environmental Externalities (January 27, 2014). CESifo Working Paper Series No. 4562. Available at SSRN

Manne A, Richels R (1992) Buying grennhouse insurance: the economic cost of carbon dioxide emission limits. MIT Press, Cambridge

Manne A, Richels R (2005) An integrated assessment model for global climate change. In: Loulou R, Waaub J-P, Zaccour G (eds) Energy environment. Springer, New York, pp 175–189

Manne AS, Mendelsohn R, Richels R (1995) MERGE: a model for evaluating regional and global effects of GHG reduction policies. Energy Policy 23:17–34

Nash J (1950) Equilibrium points in n-person games. Proc Natl Acad Sci 36:48–49

Nordhaus WD (2007) A review on the "Stern review of the economics of climate change". J Econ Lit 45(3):686–702

Nordhaus WD, Boyer J (2000) Warming the world: economics of climate change. MIT Press, Cambridge, MA

Nordhaus WD, Yang Z (1996) A regional dynamic general-equilibrium model of alternative climate change strategies. Am Econ Rev 86:741–765

Pindyck R (2013) Climate change policy: what do the models tell us? J Econ Lit 51(3): 860–872

Stern N (2006) The economics of climate change. The stern review. Cambridge University Press, Cambridge

Todorov E (2009) Efficient computation of optimal actions. Proc Natl Acad Sci 106(28): 11478–11483

Tol RS (2002a) Estimates of the damage costs of climate change. Part I: benchmark estimates. Environ Resour Econ 21:47–73

Tol RS (2002b) Estimates of the damage costs of climate change. Part II. Dynamic estimates. Environ Resour Econ 21:135–160

Tol RS (2008) The social cost of carbon: trends, outliers and catastrophes. Econ: Open-Access, Open-Assess E-J 2(25):1–22

Chapter 15
A Robust Noncooperative Meta-game for Climate Negotiation in Europe

Frédéric Babonneau, Alain Haurie, and Marc Vielle

Abstract In this paper we define and solve a 'robust game design' problem that could be used to assess the fair sharing of the abatement burden among the 28 EU countries in their coming climate negotiations. The problem consists in finding a distribution of a global 'safety emission budget' for the panning period 2010–2050, among the 28 countries in such a way to obtain a balanced relative loss of welfare (computed in percent of the discounted consumption in the reference case) when the countries supply strategically their permit endowment on a permit trading system with full banking and borrowing. We assume that the countries play a noncooperative game, where the payoffs are constituted of the gains from the terms of trade plus the gains in the permit trading and minus the abatement cost, expressed as the compensative variation of income. These payoff functions are estimated from an ensemble of numerical simulations of a detailed CGE model, GEMINI-E3 representing the economic interactions among the 28 EU countries. To deal with the uncertainty introduced by the statistical emulation technique we propose to use the concept of robust equilibrium, where the results of robust optimization are exploited in the definition of an equilibrium solution, when the payoff is subject to uncertainties. A numerical illustration is performed and an interpretation of the impact of the robustification approach on the solution of the game design problem is provided.

F. Babonneau (✉)
ORDECSYS, Place de l'Etrier 4, CH-1224 Chêne-Bougeries, Switzerland

LEURE Laboratory, Swiss Federal Institute of Technology at Lausanne (EPFL), Lausanne, Switzerland
e-mail: fbabonneau@ordecsys.com

A. Haurie
ORDECSYS, Place de l'Etrier 4, CH-1224 Chêne-Bougeries, Switzerland
e-mail: ahaurie@ordecsys.com

M. Vielle
LEURE Laboratory, Swiss Federal Institute of Technology at Lausanne (EPFL), Lausanne CH-1015, Switzerland
e-mail: marc.vielle@epfl.ch

© Springer International Publishing Switzerland 2016
F. Thuijsman, F. Wagener (eds.), *Advances in Dynamic and Evolutionary Games*, Annals of the International Society of Dynamic Games 14,
DOI 10.1007/978-3-319-28014-1_15

Keywords Game theory • Robust optimization • Meta-game • Climate negotiation

MSC Codes: 91A10, 91B50, 90C15

15.1 Introduction

In this paper we use a robust meta-game model to assess a fair burden sharing for Greenhouse Gases (GHG) abatement among the 28 European Union (EU) countries. The term 'meta-game' refers to the hierarchical structure of the game model and to the use of statistical emulation of a large-scale Computable General Equilibrium (CGE) model, GEMINI-E3 (Bernard and Vielle 2008), for the identification of the payoff functions used in the game. The term 'robust' refers to the use of the concept of robust optimization, to compute robust equilibrium, taking into account the uncertainty resulting from the statistical emulation approach. This meta-game approach has already been proposed, for a division of the whole world in a few coalitions (four or five) in Drouet et al. (2008, 2010) and more recently in Babonneau et al. (2013) and Haurie et al. (2013), where the robustification of the game has also been proposed. The originality of the present application lies in the consideration of a realistic context corresponding to the climate policy to be decided by the EU, with 28 countries as players in the game. We will also evaluate in more details the impact of robust equilibrium analysis on the solution of the burden sharing problem.

In a companion paper (Babonneau et al. 2014), the economic aspects of our approach are more fully developed and its implication for the assessment of the climate policy adopted in Europe is discussed. The present paper is more technical. Its focus is on the mathematical formulation of the robust game model, its resolution and an evaluation of the impact of robustification on the solution of the burden sharing problem that we present as a game design problem.

15.2 Fair Burden Sharing in EU Climate Policy

15.2.1 The Game Design Problem

We view the upcoming EU negotiation on climate as a 'game design problem'. More specifically one considers 2050 as the planning horizon and evaluates the global effort from Europe as a whole in terms of a global cumulative emission budget for the period 2010–2050, which we will call safety_budget. This global budget will be distributed among the 28 countries. The EU has already implemented a market for a trading emission permits that is applied only to electricity generation and energy intensive industries (Venmans 2012). We assume that the market will be extended to all sectors including households and that full banking and borrowing will be allowed. This means that the countries will have the possibility to supply strategically over time their respective global allowances on the permit market in order to optimize their payoff. The fair solution to the burden sharing problem will consist in finding a repartition of the global safety budget among the 28 countries in such a way that the payoffs received at equilibrium satisfy a Rawlsian criterion of distributive justice (Rawls 1071). This is what we call a game design problem.

15.2.2 Relevance to EU Climate Negotiation

The declared EU goal is to limit to 2 °C the global warming at the end of the century. According to Matthews et al. (2009), this goal can be translated into a limit on cumulative emissions budget of CO_2. This requires an increase of the abatement after 2020 and a worldwide reduction of 50 % of GHG emissions in 2050. In 2011 the European Commission has defined a roadmap to a low carbon economy in 2050 (European Commission 2011a). In this document the Commission reconfirmed the EU objective of reducing European GHG emissions by 80 % in 2050 compared to 1990 levels. A recent modelling analysis (European Commission 2011a) has shown that the pathway to a low carbon society requires a 40 % reduction of GHG emissions by 2030 and 60 % reduction by 2040, as displayed in Fig. 15.1. In the following analysis on the design of EU climate agreements, we constraint European countries to satisfy a global EU emissions budget for the period 2011–2050 compatible with the pathway associated to 80 % reduction by 2050. It corresponds to a safety budget of 99 Gt CO_2.

15.3 Definition of Robust Equilibrium in a *m* Player Game

The originality of our approach lies mainly in the use of a robust optimization technique in the computation of the equilibrium in permit supply among the 28 EU countries. In this section we recall the definition and characterization of robust equilibrium solutions in an *m*-player game.

15.3.1 From Nash Equilibrium to Robust Nash Equilibrium

We first recall the definition of a Nash equilibrium in a game in normal form (Nash 1950). The game is defined by:

- a set of players $j = \{1, 2, \ldots, m\}$;
- a set of strategies S_j for each player $j \in M$, where $S_j \subset \mathbb{R}^{m_j}$ is convex;
- a set of payoffs $\phi_j(\mathbf{x}) : \mathbf{x} \in \mathbf{S} = \times_{i=1}^{m} S_i \to \mathbb{R}$, where we assume that $\phi_j(\cdot)$ is continuously differentiable and concave in $x_j \in S_j$ and continuous in $\mathbf{x}^{-j} = (x_i)_{i \neq j}$.

Definition 1. A Nash equilibrium is a strategy m-tuple $\mathbf{x}^\star \in \mathbf{S}$ such that, for each player $j = 1, \ldots, m$, the following holds

$$\phi_j(\mathbf{x}^\star) = \max_{x_j \in S_j} \phi_j([\mathbf{x}^{\star -j}, x_j]), \tag{15.1}$$

where $[\mathbf{x}^{\star -j}, x_j]$ is the strategy m-tuple obtained by replacing in $\mathbf{x}^\star$ the component $x_j^\star$ by x_j.

If the strategy set of Player j is defined by a set of inequality constraints

$$S_j = \left\{ x_j \in \mathbb{R}^{m_j} : h_j(x_j) \geq 0 \right\}, \tag{15.2}$$

where $h_j(\cdot) : \mathbb{R}^{m_j} \to \mathbb{R}^{p_j}$ is continuously differentiable and concave, and under usual constraint qualification conditions, a Nash equilibrium is characterized by the following first order conditions

$$\begin{aligned} \nabla_{x_j} \phi_j(\mathbf{x}) - \nu_j^T \frac{\partial}{\partial x_j} h_j(x_j) &= 0 \\ h_j(x_j) &\geq 0 \\ \nu_j &\geq 0 \\ \nu_j^T h_j(x_j) &= 0 \\ j = 1, \ldots, m& \end{aligned}$$

Now we suppose that the payoff to Player j is rewritten $\varphi_j(\mathbf{x}, \xi)$, where φ_j is differentiable in x_j and continuous in $\mathbf{x}$ and ξ_j. Here ξ_j is an uncertain parameter which takes value in an uncertainty set $\varXi_j$ which is supposed to be closed and bounded. As proposed in Aghassi and Bertsimas (2006) we define the *robust* payoff for Player j as the worst case function

$$\phi^\rho(\mathbf{x}) = \min_{\xi_j \in \varXi_j} \varphi_j(\mathbf{x}, \xi_j). \tag{15.3}$$

Definition 2. A robust equilibrium for the game defined by the uncertain payoffs $\varphi_j(\mathbf{x}, \xi_j)$, is a Nash equilibrium for the game defined by the worst-case payoffs $\phi^\rho(\mathbf{x})$.

Assume that $\varphi_j(\mathbf{x}, \xi_j)$ has the following structure

$$\varphi_j(\mathbf{x}, \xi_j) = \underbrace{\bar{\varphi}_j(\mathbf{x})}_{\text{certain part}} + \underbrace{(P^T x)^T \xi_j}_{\text{uncertain part}},$$

where $\bar{\varphi}_j(\mathbf{x})$ is the certain part of φ_j and P is a matrix of appropriate dimension. Assume also that the uncertainty set is ellipsoidal.

$$\Xi_j = \{\xi_j \mid ||\xi_j||_2 \leq k^2\},$$

Then one can apply a result of robust optimization theory [see Theorem 1 in Babonneau and Apparigliato (2010)] to solve problem (15.3) and obtain

$$\phi^\rho(\mathbf{x}) = \bar{\varphi}_j(\mathbf{x}, \xi_j) + k||(P^T x)||_2. \tag{15.4}$$

A robust Nash equilibrium is thus characterized by the following first order conditions

$$\begin{aligned}
\frac{\partial}{\partial x_j}(\bar{\varphi}_j(\mathbf{x}, \xi_j) + k||(P^T x)||_2) - \nu_j^T \frac{\partial}{\partial x_j} h_j(x_j) &= 0 \\
h_j(x_j) &\geq 0 \\
\nu_j &\geq 0 \\
\nu_j^T h_j(x_j) &= 0 \\
j = 1, \ldots, m&
\end{aligned}$$

15.4 Mathematical Formulation of the EU Climate Game Design Problem

In this section we give a mathematical formulation of the game design problem for the EU 28 countries and we characterize the robust equilibrium solutions.

15.4.1 Variables and Parameters

Design Variables

θ_j, share of the safety emission budget given to player j.

These variables define the key element of the negotiations, namely the sharing of the safety emission budget.

Strategic Variables

$\omega_j(t)$, supply of quotas by player j during period t.

We assume that once an EU country has been given a share of the emission budget, it can supply this global amount of quotas (emission rights) on the four emissions trading markets organized in each of the four different decades of the planning horizon. These supplies are strategic variables. They influence the carbon market, determining the price of carbon, then the emission levels chosen by each country, and, finally, the transfers (buying and selling of permits) and the net surplus variations.

Secondary (Passive) Variables

These are variables that will be computed from the values given to the strategic variables. They will be used to describe the permits market functioning. The abatements realized with respect to the business as usual (BAU) scenario are the argument of the abatement cost and of the Gains from the Terms of Trade (GTT) functions that have been identified through regression analysis of a sample of GEMINI-E3 numerical simulations.

$e_j(t)$: emission level for player j in period t;
$q_j(t)$: abatement level for player j in period t;
$p(t)$: carbon price in period t;
$AC_j(t)$: abatement cost for player j in period t;
$MAC_j(t)$: marginal abatement cost for player j in period t;
$GTT_j(t)$: GTT for player j in period t;
ν_j: multiplier associated with the share of budget given to player j.

Parameters

safety_budget: global safety emission budget;
$\mathrm{bce}_j(t)$: BAU emissions for player j in period t;
ny(t): number of years in period t;
n(t): number of years in time interval $[1, t]$;
$\alpha_j^0(t)$, $\alpha_j^1(t)$, $\alpha_j^2(t)$, $\alpha_j^3(t)$, $\alpha_j^4(t)$: coefficients in the abatement cost function;
$\mu_j^0(t)$, $\mu_j^1(t)$: coefficients in the gain from the terms of trade function;
β: discount factor;
hc_j: discounted household consumption in BAU over the planning horizon.

15.4.2 Payoffs and Constraints

Payoffs for the Game of Quotas Supply. We assume that the countries try to minimize its discounted sum of surplus losses, denoted $-W_j$ while taking into account the strategic actions taken by the other players.

$$-W_j = \sum_t \beta^{\mathrm{n(t)}} \mathrm{ny(t)} \left\{ AC_j(t) - p(t)(\omega_j(t) - e_j(t)) - GTT_j(t) \right\}, \qquad (15.5)$$

is the discounted sum of the abatement cost, measured as the Compensative Variation of Income (CVI), minus the GTT and the revenue from the permit trading (can be negative). Equivalently we will define the payoffs W_j in terms of surplus gains instead of losses.

Optimality Criterion for the Game Design Problem. At the upper level where one negotiates the sharing of the safety emissions budget, one will apply a criterion of fairness inspired from the Rawlsian theory of justice:

$$z = \max_{\boldsymbol{\theta}} \min_{j} \frac{W_j^\star}{\mathrm{hc}_j}, \tag{15.6}$$

where $W_j^\star$ is the equilibrium payoff for the game designed by the choice of $\boldsymbol{\theta}$ and hc_j is the discounted sum of consumption in the BAU case. So we select the sharing which, in the Nash equilibrium solution of the game of quotas supply, maximizes the worst relative surplus gain among the players.

Constraints and Functions. They link the passive variables to the strategic variables, define the cost and profit functions, limit the choices for the strategic variables.

Shares of safety budget: The total supply of quotas by each player is equal to its share of the safety budget:

$$\sum_t \omega_j(t) = \theta_j \,\text{safety\_budget}. \tag{15.7}$$

Lagrangian multipliers ν_j are associated to these constraints.

Price of carbon equal marginal abatement cost: In a competitive emission permits market, each player will abate at a level where the price of permit equals the marginal abatement cost:

$$p(t) = MAC_j(t), \forall t, j. \tag{15.8}$$

Permit market clears: In this market, the price is set at such a level that the total emission equals the total supply of quotas:

$$\sum_j \omega_j(t) = \sum_j e_j(t), \forall t. \tag{15.9}$$

Define emissions from abatements: One must compute abatement level to evaluate abatement costs:

$$e_j(t) + q_j(t) = \mathrm{bce}_j(t). \tag{15.10}$$

Abatement cost: The abatement cost is a polynomial of degree 4[1] in the abatement variable:

[1]Note that, as the second derivative of the AC appears in the mathematical game formulation, we have thus imposed a convexity constraint on this second derivative in the regression model in order to ensure the convexity of the overall problem. Moreover, polynomial forms of lower degree have been tested but resulting in worse estimation quality.

$$AC_j(t) = \alpha_j^1(t)\, q_j(t) + \alpha_j^2\, q_j(t)^2 + \alpha_j^3(t)\, q_j(t)^3 + \alpha_j^4(t)\, q_j(t)^4. \tag{15.11}$$

Marginal abatement cost: The marginal abatement cost is obtained through derivation of the abatement cost:

$$MAC_j(t) = \alpha_j^1(t) + 2\,\alpha_j^2(t)\, q_j(t) + 3\,\alpha_j^3(t)\, q_j(t)^2 + 4\,\alpha_j^4(t)\, q_j(t)^3. \tag{15.12}$$

Derivative of marginal abatement cost: One also needs to compute the derivative of the marginal cost function:

$$DMAC_j(t) = 2\,\alpha_j^2(t) + 6\,\alpha_j^3(t)\, q_j(t) + 12\,\alpha_j^4(t)\, q_j(t)^2. \tag{15.13}$$

Gains from the terms of trade: The GTT are expressed as a linear function of the sum of the abatements decided by all the players:

$$GTT_j(t) = \mu_j(t) \sum_i q_i(t). \tag{15.14}$$

15.4.3 Necessary Conditions for Equilibrium

Derivative of carbon price: The derivative of the market carbon price w.r.t. the total supply $\omega(t)$ is given by [see Helm (2003)]:

$$DP(t) = \frac{-1}{\sum_j \frac{1}{DMAC_j(t)}}. \tag{15.15}$$

Pseudo-gradient of payoffs: The pseudo-gradient of the payoffs W_j w.r.t. the strategic variables $\omega_j(t)$ is given by

$$PSGRAD_j(t) = -\beta^{\mathrm{n(t)}} \mathrm{ny(t)} \left\{MAC_j(t) - DP(t)\,(\omega_j(t) - e_j(t)) - \mu_j^1(t)\right\} + \nu_j. \tag{15.16}$$

First Order Conditions for a Nash Equilibrium. To compute a Nash equilibrium one has to solve the following equations:

$$\begin{aligned} \nu_j &\geq 0 \\ \theta_j \,\text{safety\_budget} - \sum_\tau \omega_j(\tau) &\geq 0 \\ \nu_j\,(\theta_j \,\text{safety\_budget} - \sum_\tau \omega_j(\tau)) &= 0 \\ &\forall j \end{aligned}$$

$$
\begin{aligned}
-PSGRAD_j(t) &\geq 0 \\
\omega_j(t) &\geq 0 \\
\omega_j(t)\, PSGRAD_j(t) &= 0 \\
&\qquad \forall j, \forall t.
\end{aligned}
$$

15.4.4 Robustifying the Game Models

We now introduce robustness in the game analysis by applying robust optimization techniques introduced in Ben-Tal et al. (2009) to approximation error of marginal abatement cost function.

The reward function used for each period and each player is uncertain, since it is obtained through statistical emulation of a large scale CGE model, GEMINI-E3. In particular, as it will be shown in Sect. 15.5, we estimate in this way the avoided abatement cost. In order to extract robust predictions from the game meta-model we must robustify the marginal abatement cost (15.12) in which coefficients $\alpha_j^i(t)$ are estimated using a linear regression techniques as explained in Babonneau et al. (2013). Let $\bar{\alpha}_j^i(t)$ be this estimate and $\hat{\alpha}_j^i(t)$ be the estimate error, we describe the uncertain coefficients as linear functions of an underlying random factor ξ_j^t

$$\alpha_j^i(t) = \bar{\alpha}_j^i(t) + \xi_{ij}^t \hat{\alpha}_j^i(t)$$

Let us consider an uncertainty set defined as follows

$$\Xi_j^t = \{\xi : \sum_{i=1}^{4} |\xi_{ij}^t|^2 \leq k^2\}.$$

Using robust optimization techniques sketched in Sect. 15.3, the worst case of the marginal abatement cost function is given by

$$\overline{MAC}_j(t) = \sum_{i=1}^{4} i\bar{\alpha}_j^i(t)\,(q_j^t)^{i-1} + k\sqrt{\sum_{i=1}^{4} (i\hat{\alpha}_j^i(t)\,(q_j^t)^{i-1})^2}.$$

Probabilistic Interpretation. The second component above corresponds to a safety factor ensuring that the marginal abatement cost for all realizations of ξ within the uncertainty set Ξ_j^t will be lower than the worst case $\overline{MAC}_j(t)$. One can derive a satisfaction probability for the marginal abatement cost to be lower than $\overline{MAC}_j(t)$ for any realizations of ξ that depends on the radius k of the uncertainty set. This probabilistic result is given in Theorem 1 below. This factor k plays a crucial role as the larger its value, the greater the number of realizations ξ that are considered.

Theorem 1 (From Ben-Tal et al. (2009) with Proof). *Let* ξ_i, $i = 1, \ldots, n$ *be independent random variables with values in interval* $[-1, 1]$ *and with average zero:* $E(\xi_i) = 0$*. If* z_i, $i = 1, \ldots, n$*, are deterministic coefficients, we have for all* $k \geq 0$

$$Prob \left\{ \xi \mid \sum_{i=1}^{n} z_i \xi_i > k \sqrt{\sum_{i=1}^{n} z_i^2} \right\} \leq \exp(-\frac{k^2}{2.5}).$$

Remark 1. The probabilistic interpretation given in Theorem 1 depends on the support of the random variables, their expectation and their assumed independence. No other assumption concerning the form of the probability distributions is made.

In the numerical experiment presented in Sect. 15.6, we use $k = 3$. This leads to a 97 % probability of satisfaction.

15.5 Implementation Issues

In this section we calibrate the robust dynamic game model, using the CGE GEMINI-E3 as the provider of the data used in the estimation of the abatement cost functions for each EU country. The GAMS implementation of the robust formulation of the game is given in Appendix.

15.5.1 Estimation of the Abatement Cost Functions

GEMINI-E3, a CGE Model. GEMINI-E3 (Bernard and Vielle 2008) is a multi-country, multi-sector, recursive CGE model comparable to the other CGE models (EPPA, ENV-Linkage, etc.) built and implemented by other modeling teams and institutions, and sharing the same long experience in the design of this class of economic models. GEMINI-E3 has been extensively used to derive total costs and benefits of various energy and climate policies. The GEMINI-E3 model is now built on a comprehensive energy-economy dataset, the GTAP-8 database (Narayanan et al. 2012). This database incorporates a consistent representation of energy markets in physical units, social accounting matrices for each individualized country/region, and the whole set of bilateral trade flows.

A Business as Usual Scenario. We build a BAU scenario on the period 2007–2050 with yearly timesteps. Assumptions on population and GDP are based on the recent joint work of the Economic Policy Committee and the European Commission (DG ECFIN) published in 2011 (European Commission 2011b). They suppose that European GDP will grow by 1.6 % per year over the period 2010–2050. Evolution of energy prices are based on assumptions on the *current policies* scenario of the *World Energy Outlook 2013* of the International Energy Agency (International Energy Agency 2013). The oil price is assumed to reach 162$ in 2050, the price of imported

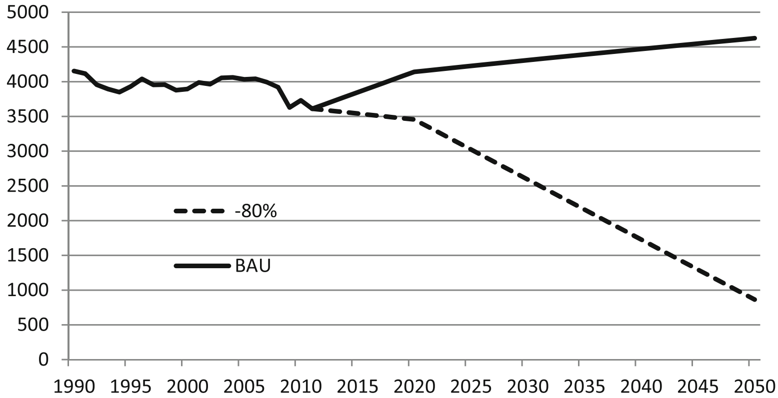

Fig. 15.1 EU CO_2 emissions in the BAU scenario (1990–2010: historical, 2011–2050: GEMINI-E3 BAU scenario) and climate target in Mt CO_2

gas in Europe is equal to 15.6$ per Mbtu in 2050, and the price of steam coal imported in OECD countries reaches 125$ per ton in 2050.

Note that, in this BAU scenario, no climate policy is implemented, since it will serve to evaluate the burden for each participating country of implementing the European climate policy, considering 2050 target as well as existing 2020 objectives. Associated CO_2 emissions computed by GEMINI-E3 are presented in Fig. 15.1. In 2050, total European CO_2 emissions reaches 4625 MtCO_2 corresponding to an annual growth rate of 0.5 %. Our BAU is consistent with the "the no-policy baseline scenario" performed within the EMF28 project (Knopf et al. 2013) where most of the models suggest a more modest increase of CO_2 emissions. Our emissions will generate a cumulative emissions budget of 173 Gt CO_2 over the period 2011–2050.

Statistical Analysis of a Sample of GEMINI-E3 Numerical Simulations. We apply regression analysis to identify the payoff functions of a game where the strategic variables are the quotas supplied on an EU emissions trading scheme by the different regions, at different periods. The statistical analysis is based on a sample of 200 numerical simulations of different possible European climate policy scenarios performed with GEMINI-E3. In each scenario, we assume that a carbon tax is implemented at the European level without emissions trading. We suppose that only carbon emissions are taxed. We compute for each group of countries:

- The abatement level relative to the BAU emissions ($\bar{e}_j^t$) expressed in million ton of carbon; The abatement is thus defined by $\bar{e}_j^t - e_j^t$
- The welfare cost measured by the households' surplus, and represented by the CVI expressed in US $ (Bernard and Vielle 2003);
- The GTT representing the spill-over effects due to changes in international prices. In a climate change policy these GTT come mainly from the drop in fossil energy prices due to the decrease of world energy demand. The GTT are expressed in US $.

By subtracting the GTT from the surplus we obtain the deadweight loss of taxation i.e. the domestic cost that would occur in a closed economy and which only depends on the abatement done within the country. The GTT represents the imported cost: negative for energy exporting countries such as OPEC and positive for net energy importing countries like Europe and Japan (Böhringer and Rutherford 2002). This imported cost/benefit is function of the European carbon abatement.

Using regression analysis, we estimate the parameters $\alpha_j^i(t)$ of the abatement cost function $AC_j(t)$ given in (15.11) for player j and period t. The time periods (t) are 2020, 2030, 2040, 2050 with 10 years for each period.

The GTT of player j is assumed to be a linear function of the global abatement in a given period

$$GTT_j(t) = \mu_j(t) \sum_i (\bar{e}_i^t - e_i^t). \tag{15.17}$$

These estimations serve to define the payoff functions introduced in the game of Sect. 15.4.

15.6 Numerical Illustration

In this section we illustrate the potential impact of uncertain abatement costs when one designs solutions using deterministic models. We first compute a deterministic balanced equilibrium corresponding to an equilibrium solution where the welfare losses, expressed in percent of total discounted consumption, are almost equal among the 28 EU countries. This solution is simply obtained by using the deterministic version of the game and by adjusting properly the weights θ_j until a balanced equilibrium is reached. The balanced burden sharing and the welfare losses are given in Table 15.1 in columns 2 and 3, respectively. We observe a total EU welfare loss of 1.18 % of total discounted household consumption. The quotas are concentrated among a limited number of countries, the top 5 represents 61 % of the budget and the top 10 82 %. We find in the top 5 the largest European economy: Germany, United Kingdom, France, Italy and Spain.

We then evaluate the welfare losses associated to this deterministic balanced solution (column 2 of Table 15.1) using the robust version of the game that takes into consideration the errors of meta-modelling observed in the calibration of the coefficients $\alpha_j^i(t)$ in the abatement cost functions. Let $\sigma_j^i(t)$ be standard deviation estimators computed from linear regression techniques associated to coefficients $\alpha_j^i(t)$, we assume variabilities $\hat{\alpha}_j^i(t) = 2 \cdot \sigma_j^i(t)$. Column 4 of Table 15.1 shows the results for $k = 3$. This corresponds to a confidence interval of 96 %. First and as expected, the robust model yields a larger EU welfare loss (2.23 % of total discounted household consumption). More importantly the deterministic burden sharing does not anymore lead to a balanced solution (column 4 of Table 15.1). Welfare losses now range from −0.25 to 6.41, with generally an increase of the

Table 15.1 Deterministic vs. robust balanced burden sharing and associated deterministic and robust welfare losses (in % of total discounted household consumption)

	Burden sharing	Welfare losses		Burden sharing	Welfare losses
	Deterministic	Deterministic	Robust	Robust	Robust
Austria	1.85	1.04	2.20	1.85	2.20
Belgium	3.27	1.07	2.18	3.27	2.18
Bulgaria	0.83	1.07	1.55	0.82	2.26
Cyprus	0.42	0.98	6.23	0.46	2.28
Czech Republic	1.38	1.20	5.97	1.63	2.26
Germany	13.81	1.25	3.76	16.07	2.23
Denmark	1.85	1.05	2.46	1.88	2.22
Estonia	0.40	1.09	2.69	0.40	2.28
Finland	1.50	1.12	1.38	1.42	2.12
France	12.15	1.20	2.01	11.90	2.21
United Kingdom	14.82	1.20	1.27	13.30	2.26
Greece	5.05	1.19	6.41	5.75	2.28
Croatia	0.85	1.13	2.23	0.85	2.23
Hungary	1.30	1.01	1.55	1.25	2.21
Ireland	1.43	1.12	2.29	1.43	2.29
Italy	11.72	1.20	1.87	11.30	2.29
Latvia	0.30	1.15	3.54	0.32	2.23
Lithuania	0.30	1.23	3.66	0.33	2.26
Luxembourg	0.56	0.99	4.13	0.60	2.21
Malta	0.13	0.94	5.40	0.15	2.15
Netherlands	4.40	1.01	2.62	4.50	2.28
Poland	6.22	1.14	1.00	5.95	2.21
Portugal	1.50	1.06	1.92	1.47	2.22
Romania	1.93	1.12	−0.25	1.71	2.22
Spain	8.95	1.21	1.45	8.50	2.13
Slovak Republic	0.84	1.09	0.98	0.79	2.21
Slovenia	0.46	1.15	2.46	0.46	2.27
Sweden	1.79	1.25	1.43	1.65	2.18
EU-28	100.00	1.18	2.23	100.00	2.23

welfare loss. The countries which experience the worse degradation are Cyprus, Greece, Malta, Czech Republic and Luxembourg. In contrary, only three countries are in a better situation within the robust game: Slovak Republic, Poland and Romania. Regarding the major economic areas, only Germany is significantly impacted when considering the robust version. In the last two columns of Table 15.1 we display the robust balanced solution (burden sharing and welfare losses) computed with the robust formulation. Of course the quotas are redistributed to the benefit of countries whose welfare losses increase significantly, for example the

Table 15.2 CO_2 prices in \$ per t CO_2

	Deterministic	Robust
2020	240	360
2030	391	586
2040	637	955
2050	1037	1554

budget allocated to Germany rises from 13.8 % to 16.1 %. But the hierarchy between countries does not change significantly and we find in the tops 5 and 10 the same countries.

Let us now observe CO_2 prices associated to deterministic and robust solutions and given in Table 15.2. For the deterministic solution, the CO_2 price reaches 1037 US \$ in 2050. This price is in the range of the carbon prices found in the AMPERE (Capros et al. 2014) (243–824 €/tCO_2) and EMF28 (Weyant et al. 2013; Knopf et al. 2013) (240–1127 €/tCO_2) projects. When considering the uncertainties related to the abatement cost functions the price in 2050 increases by 50 % and is equal to 1554 US \$.

15.7 Conclusion

In this paper we have shown how a meta-game approach could provide an assessment of the fair sharing of the abatement burden among the 28 countries negotiating the EU climate policy with 2050 as an horizon. The game structure exploits the existence of a permit trading market for the whole of EU, with full banking and borrowing possibility. The countries payoff functions have been identified through statistical emulation, based on an ensemble of 200 scenarios, of a detailed multi-country CGE, GEMINI-E3. To cope with the uncertainty which is inherent of this statistical emulation approach, we have proposed to use the notion of robust equilibrium as a solution concept. We show that the 80 % reduction target can be achieved with a significant but acceptable welfare losses at EU level. We see that the robustification of the equilibrium concept has an important effect on the fair sharing of the global safety emission budget and can significantly increase the cost of the climate policy. Our results are in line with those computed from large-scale economic models (Capros et al. 2014; Weyant et al. 2013; Knopf et al. 2013). The proposed meta-Game approach, which results in a small and tractable robust model could therefore be a useful tool to help in the definition of the next round of climate negotiations.

Acknowledgements The research leading to these results has received funding from Qatar National Research Fund under Grant Agreement no 6-1035-512.

Appendix

GAMS Code of the Robust Game

We report below the GAMS implementation code of the robust formulation to the game. This model is solved with the PATH solver (Ferris and Munson 2000).

```
sets
J   PLAYERS
T   PERIODS 2020 TO 2050
;

Table
EX(J,T)   EXCHANGE RATES
GTT(J,T) GAINS FROM TERMS OF TRADE LINEAR TERM
EB(J,T)   BAU emissions
A0(J,T)   COEFF. CONST. MAC FUNCTION
A1(J,T)   COEFF. LIN. MAC FUNCTION
A2(J,T)   COEFF. QUAD. MAC FUNCTION
A3(J,T)   COEFF. CUB. MAC FUNCTION
A1_var(J,T) VARIABILITY OF COEFF. LIN. MAC FUNCTION
A2_var(J,T) VARIABILITY OF COEFF. QUAD. MAC FUNCTION
A3_var(J,T) VARIABILITY OF COEFF. CUB. MAC FUNCTION
;

scalar
BUD  GLOBAL EMISSION BUDGET
beta   DISCOUNT FACTOR
;

parameter
BS(J) SHARES OF EM. BUDGET
;

positive variable
a(T,J) ALLOWANCES
e(T,J) EMISSION LEVELS
q(T,J) ABATEMENT LEVELS
P(T)   PERMIT PRICES
nu(J)  MULTIPLIER ALLOWANCE
;

variable
DAcost(T,J)    MARGINAL ABATEMENT COSTS
DDAcost(T,J)   SECOND DERIVATIVE ABATEMENT COST
```

```
totA(T)         TOTAL ALLOWANCES
totE(T)         TOTAL EMISSIONS
cE              CUMULATIVE EMISSIONS
dP(T)           DIFF PRICE
dE(T,J)         DIFF EMISS
TR(T,J)         NET TRANSFERS
;

equations

* CHECK THAT THE SOLUTION USES  THE BUDGET SHARE
BcumA(j).. BUD*BS(J)- (10*sum(T, a(T,J))) =g= 0;

* DEFINES TOTAL ALLOWANCES AT T
EtotA(T).. totA(T) -  sum(J,a(T,J)) =e= 0;

* DEFINES TOTAL EMISSIONS
EtotE(T).. totE(T) -  sum(J,e(T,J)) =e= 0;

* DEFINES EMISSIONS FROM ABATEMENT AT T
EQe(T,J).. -EB(J,T)+ e(T,J) + q(T,J)  =e= 0;

* DEFINES TOTAL EMISSIONS
EQce..   cE  =e= 10*sum(T, totE(T));

* DEFINES MAC
EQDAcost(T,J).. DAcost(T,J) - (A1(J,T)*q(T,J)+
    A2(J,T)*q(T,J)**2+A3(J,T)*q(T,J)**3)/EX(J,T)
   - k2*sqrt(abs(A1_var(J,T)*q(T,J))**2 +
   abs(A2_var(J,T)*q(T,J)**2)**2 +abs(
   A3_var(J,T)*q(T,J)**3)**2)/EX(J,T) =e= 0;

* DEFINES MINUS DERIVATIVE OF MAC
EDDAcost(T,J)..   DDAcost(T,J)+(A1(J,T)+2*A2(J,T)
   *q(T,J)+3*A3(J,T)*q(T,J)**2)/EX(J,T) + k2*(2*
   abs(A1_var(J,T))**2*q(T,J)+4*abs(A2_var(J,T))**2
   *q(T,J)**3 + 6*abs(A3_var(J,T))**2*q(T,J)**5)
   / (2*EX(J,T)*sqrt( abs(A1_var(J,T)*q(T,J))**2
   +abs(A2_var(J,T)*q(T,J)**2)**2
   +abs(A3_var(J,T)*q(T,J)**3)**2)) =e= 0;

* DEFINES DERIVATIVE OF MARKET PRICE WRT ALLOWANCE
EdP(T)..  dP(T) - 1/sum(J,1/DDAcost(T,J)) =e= 0;

* DEFINES DERIVATIVE OF EMISSION WRT ALLOWANCE
```

```
EdE(T,J)..  dE(T,J) -  1/sum(I,DDAcost(T,J)
   /DDAcost(T,I)) =e= 0;

* DEFINES PSEUDO-GRADIENT OF PAYOFFS W.R.T.
* ALLOWANCES, TAKING INTO ACCOUNT EFFECTS ON PRICES.
PSGRAD(T,J).. -(1+beta)**(-10*(ord(T)-1))
   *(DAcost(T,J)
     + dP(T)*(a(T,J)-e(T,J)) - GTT(J,T))
   + Nu(J)
     * (sqrt(var*sum(I,sum(TI,(10*a(TI,I))**2)))
        + a(T,J)*k*var)
   =e= 0;

* MARKET CLEARS (TOTAL EMISSIONS EQUAL TOTAL
* ALLOWANCES AT T)
MARKETC(T).. totA(T)- totE(T) =e= 0;

* PRICE IS EQUAL TO MAC
MAXPRO(T,J)..  DAcost(T,J) - P(T) =e= 0;

* TRANSFERS
TRANS(T,J).. TR(T,J)-P(T)*(a(T,J)-e(T,J)) =e= 0;


model robust-game
/
BcumA.Nu,
EtotA.totA,
EtotE.totE,
EQe.q,
EQce.cE,
EQDAcost.DAcost,
EDDAcost,
EdP.dP,
EdE.dE,
PSGRAD.a,
MARKETC.P,
MAXPRO.e,
TRANS.TR
/;

option mcp=path;
solve robust-game using mcp;
```

References

Aghassi M, Bertsimas D (2006) Robust game theory. Math Program Ser B 107:231–273

Babonneau F, Haurie A, Vielle M (2013) A robust meta-game for climate negotiations. Comput Manag Sci 10(4):299–329

Babonneau F, Haurie A, Vielle M. Assessment of Balanced Burden Sharing in the 2050 EU Climate/Energy Roadmap: A Metamodeling Approach, Climatic Change, available online, DOI 10.1007/s10584-015-1540-x

Babonneau F, Vial J-P, Apparigliato R (2010). Robust optimization for environmental and energy planning. In: Filar JA, Haurie A (eds) Uncertainty and environmental decision making. Springer, Berlin

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization. Princeton University Press, Princeton

Bernard A, Vielle M (2003) Measuring the welfare cost of climate change policies: a comparative assessment based on the computable general equilibrium model GEMINI-E3. Environ Model Assess 8:199–217

Bernard A, Vielle M (2008) GEMINI-E3, a general equilibrium model of international national interactions between economy, energy and the environment. Comput Manag Sci 5:173–206

Böhringer C, Rutherford TF (2002) Carbon abatement and international spillovers. Environ Resour Econ 22:391–417

Capros P, Paroussos L, Fragkos P, Tsani S, Boitier B, Wagner F, Busch S, Resch G, Blesl M, Bollen J (2014). Description of models and scenarios used to assess European decarbonisation pathways. Energy Strategy Rev 2(3–4):220–230

Drouet L, Haurie A, Moresino F, Vial J-P, Vielle M, Viguier L (2008) An oracle based method to compute a coupled equilibrium in a model of international climate policy. Comput Manag Sci (Special issue “Managing Energy and the Environment”) 5(1):67–94

Drouet L, Haurie A, Vial J-P, Vielle M (2010) A game of international climate policy solved by a homogenous oracle-based method for variational inequalities. In: Breton M, Szajowski K (eds) Advances in dynamic games. Annals of the International Society of Dynamic Games, vol 11. Birkhäuser, Basel, pp 469–488

European Commission (2011a). COM/2011/112. A Roadmap for moving to a competitive low carbon economy in 2050. Communication from the Commission to the European Parliament, the Council, the European Economic and Social committee and the committee of the Regions, European Commission, Brussels, Belgium

European Commission (2011b). The 2012 ageing report: underlying assumptions and projection methodologies, European Economy 4, European Commission, Brussels, Belgium

Ferris M, Munson T (2000) Complementarity problems in GAMS and the PATH solver. J Econ Dyn Control 24:165–188

Haurie A, Babonneau F, Edwards N, Holden PB, Kanudia A, Labriet M, Pizzileo B, Vielle M (2013). Fairness in climate negotiations: a meta-game analysis based on community integrated assessment. In: Semmler W, Bernard L (eds.) Handbook on the macroeconomics of climate change. Oxford University Press, Oxford (to appear)

Helm C (2003) International emissions trading with endogenous allowance choices. J Public Econ 87:2737–2747

Energy Agency (2013) World energy outlook, IEA annual report

Knopf B, Chen Y-HH, De Cian E, Förster H, Kanudia A, Karkatsouli I, Keppo I, Koljonen T, Schumacher K, Van Vuuren DP (2013) Beyond 2020 - startegies and costs for transforming the European energy system. Clim Change Econ, 4(Supp. 1), 38 pages

Matthews HD, Gillett NP, Stott PA, Zickfeld K (2009) The proportionality of global warming to cumulative carbon emissions. Nature 459:829–832

Narayanan B, Aguiar A, McDougall R (eds) (2012) Global trade, assistance, and production: the GTAP 8 Data Base. Center for Global Trade Analysis, Purdue University

Nash J (1950) Equilibrium points in n-person games. Proc Natl Acad Sci 36:48–49

Rawls J (1071) A theory of justice. Harvard University Press, Cambridge, MA

Venmans F (2012) A literature-based multi-criteria evaluation of the EU ETS. Renew Sustain Energ Rev 16(8):5493–5510

Weyant J, Knopf B, de Cian E, Keppo I, van Vuuren D (2013) Introduction to the EMF28 Study on scenarios for transforming the European energy system. Clim Change Econ, 4(Supp. 1), 3 pages.